“Time Flies” Multisensory Processing by Circadian Clocks in *Drosophila melanogaster*

Ross EF Harper

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

CoMPLEX

University College London

August 25, 2017

I, Ross EF Harper, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Periodic changes of environmental signals are sufficient to synchronise circadian rhythms across species. Circadian time, then, is a concept tethered to a diverse spread of different sensory modalities. In spite of this fact, circadian systems have historically been studied in a unimodal fashion – investigating the processing of singular cues, while keeping others constant. My research sought to challenge this dogma via exploration of multisensory cue combination in the circadian clock of *Drosophila melanogaster*.

Systematic behavioural analysis in wild type flies showed that misalignments between light and temperature (two potent environmental cues) produced abnormal profiles of circadian locomotor activity. Further molecular investigation revealed this behavioural disruption was associated with a breakdown of molecular rhythms in central clock neurons. Both the behavioural and molecular phenotypes observed during sensory conflict depended on the circadian photoreceptor, *cryptochrome*.

Outside the central clock network, the circadian system of fruit flies forms an extensive network of peripheral oscillators. A *luciferase* reporter assay showed that photic signals play a more prominent role in peripheral clocks, compared to the core clock network in the brain. Here, molecular rhythms displayed continued light preference during sensory conflict, which again depended on *cryptochrome*.

To further explore the implications of multisensory processing, with particular focus on the blurred boundary between clock input and output, circadian gene expression was evaluated in the 'Johnston's Organ' – a key mechanosensory apparatus in *Drosophila*. These preliminary data suggest the existence of a previously unidentified peripheral clock in the fruit fly ear.

Finally, a statistical model of the circadian clock was developed using a novel graphical architecture based on the hidden Markov model framework. This model was capable of inferring the phase of an underlying clock from both simulated and experimental locomotor datasets. More broadly, learning the parameters of this model from the data produced a probabilistic representation of the system, including its phase response dynamics.

Acknowledgements

Thank you, first and foremost, to my three wonderful supervisors: Joerg Albert, Ralf Stanewsky, and Peter Dayan. You have been a constant source of guidance and support, without which I simply would not be here authoring these acknowledgements.

To my examiners, thank you for the careful and considered reading of this thesis. I appreciate the time you have given me, and I look forward to our discussions in the viva.

I've been lucky enough to work with a number of fantastic people during my PhD. Having spent time in the department of Cell & Developmental Biology, UCL Ear Institute, and the Gatsby Computational Neuroscience Unit, I've had the pleasure of getting to know a diverse mix of scientists. I cannot possibly begin to name them all.

To the members of the Albert and Stanewsky labs, past and present, thank you for your technical assistance, discussions, and friendship over the last three years. Marta Andrés, Nicholas Boyd-Gibbins, Adam Bradlaugh, Chenghao Chen, Min Chen, Xiaoli He, Assel Kashkenbayeva, Ryan Kavlie, Alyona Keder, Maite Ogueta, Sanne Roessingh, Jason Somers (the neurobiological Batman to my Robin), Camille Tardieu, and Matthew Topping, I am grateful to have worked alongside such brilliant minds and inspiring people, and I look forward to our continued relationship in the years to come.

To the members of the Gatsby Unit, rarely is such a welcoming and neighbourly community populated with such fierce intellect. Once again, I am grateful to my supervisor, Peter Dayan, for having cultivated and maintained this truly unique

environment. My time with you all has been a continuing source of enjoyment, and your thought-provoking conversations have made me a better scientist. Notable mention must go to Joëlle Ackermann, Sanjeevan Ahilan, Alexander Antrobus, and Susan Fischer, for making my time here that bit more special.

Finally, I would like to thank my family – your contribution to my research can be found on every page of this thesis. To my father, David, you are one of the greatest men I know, and it is a sincere desire to make you proud that has led me down this route. For this, I cannot thank you enough. To my mother, Lorraine, you are a rock, and a constant source of emotional support to me and the family. One cannot fathom where exactly you find the strength. To my older sister, Laura, you have always kept an eye on me, often without sufficient gratitude on my part. I have no doubt that your patience and care in my formative years made me the inquisitive and confident person I am today. (You only have yourself to blame). Lastly, to my younger brother, Edmund, you have grown into one of my best friends, and always provided a fun escape from the trials and tribulations of PhD life. I'm very proud of the person you have become, and I look forward to joining you on the next step.

I would not be the person I am today without the many people who have affected and influenced me over the years. The work presented in this thesis is my own, but it is a collaboration in many ways.

Contents

List of Figures

List of Tables

Chapter 1

Background: *Drosophila* Chronobiology

Few concepts have garnered such profound attention throughout history. Ancient cultures sought to measure it, natural philosophers and post-enlightenment scientists struggled to define it, and a popular trend in physics today is to deny its existence altogether. Yet the fact remains that almost every living thing on Earth evolved in a 24-hour periodic environment. Time, therefore, is of paramount importance to biology. While spatial organisation gives form, it is the temporal structure of biology that provides relevance.

1.1 A Brief History of Biological Time

1.1.1 Then

When the sun rises, we are not surprised. We *expect* it to do so. This inductive reasoning comes from the simple fact that the sun has always behaved in this way; indeed with such reliability as to form a pillar of certainty against which confidence itself is casually measured – 'as sure as the sun rises'.

Certainty around the solar cycles is not a modern idea. When dawn came, over two-thousand years ago in ancient Macedonia, Androsthenes of Thasos, an admiral of Alexander the Great, was as unsurprised as we are today. Instead, he made an astute observation. He noticed another more subtle daily cycle in the world around him. On the island of Tylos (now Bahrain), Androsthenes recorded the daily

opening and closing of the leaves of the tamarind tree, *Tamarindus indicus* (Bretzl, 1903). Clearly, it would seem, 24-hour cycling was present in biological processes, as well as astronomical ones. Romantically commenting that the tree appeared to be 'worshipping the sun', Androsthenes is today widely regarded as the first to document what we now refer to as a 'circadian' behaviour.

This early observation, however, lay dormant until 1729, when the French astronomer Jean-Jacques d'Ortous de Mairan demonstrated that daily cycles of leaf opening and closing in *Mimosa pudica* persisted even while the plant was kept in constant darkness (i.e. in the absence of any periodic environmental stimuli; De Mairan 1729). This exposed a fundamental truth: that 24-hour rhythms were an endogenous property of the plant itself. Unfortunately, intellectual inertia within the scientific community stalled de Mairan's revelation; the general consensus being that his results (and those of subsequent rogue academics) could be explained by poor experimental set-up. Under the erroneous justification that periodic environmental stimuli were not truly absent, rhythmic leaf opening was deemed a more simple stimulus-response behaviour, rather than an innate property of the organism.

It was not until the mid-twentieth century that the mass of accumulating evidence was sufficient to persuade the critics, and a new field of chronobiology began to take form under the guidance of those we now consider fathers of the science: Jurgen Aschoff, Erwin Bünning, and Colin S. Pittendrigh. A new-found interest in the intrinsic timing of biological systems spread throughout the academic community, with a symbolic inauguration in 1959 when the term 'circadian' was coined to refer to such 24-hour rhythms (Halberg et al., 2003). From the Latin, *circa* (about), and *diem* (day), we now attach more rigorous definitions to these biological processes (Table 1.1), along with an accompanying scientific lexicon (Table 1.2).

1.1.2 Now

Time has certainly taken a prolonged route to its current place in modern-day biology, somewhat humorously contrasting its fundamentality to all living things. The Earth was formed nearly 4.6 billion years ago when a cloud of cosmic debris collapsed under its own gravity, generating an angular momentum that persists today.

Criteria	Description
24-hour endogenous rhythms	In the absence of any periodic external input, the behaviour must cycle on a period of around 24 hours.
Entrainment to environmental rhythms	The rhythm can be reset by exposure to external stimuli such as light and heat. This distinguishes circadian rhythms from other behaviours that can oscillate on a roughly 24-hour period but are independent of environmental conditions (such as cell division, which can occur at a daily rate).
Temperature compensation	The rhythm must maintain roughly the same period over a range of physiological temperatures. Differences in thermal energy are known to affect the kinetics of molecular interactions inside cells. However, to serve as a practical timing mechanism, changes in baseline temperature must not affect the rate of circadian oscillations. What makes this unresponsiveness to baseline temperature changes particularly interesting is the simultaneous sensitivity of many clocks to daily cycles of warmth and cold (see Section 1.2).

Table 1.1: Criteria for circadian rhythms

Consequently, our planet has, and will continue to, spin on its longitudinal axis at a rate of nearly one revolution every 24 hours. This produces an inherent periodicity in earthly habitats. Light-dark, warm-cold, presence-absence of food and predators; the surrounding world is a cyclic entity. Taking into account the theory of natural selection, it seems clear that extremely early on in the history of life, there has been an almost ubiquitous evolutionary pressure to sense time, and in doing so, optimise biology for a rhythmically changing world.

Circadian rhythms have now been identified in organisms of all phyla. From single celled prokaryotes, to plants, and throughout the animal kingdom, circadian timing mechanisms underpin biology. Moreover, those few species that prove exceptions to this rule, tend to do so in a predictable manner. *P. andruzzii*, a subterranean cavefish that lives in constant darkness and temperature, has been shown to exhibit infradian (longer than a day) rhythms of up to 47 hours (Cavallari et al., 2011). This unusually long period is not as surprising as it first seems, with the internal oscillations instead well-timed to the cyclic availability of a food source. Cu-

Term	Description
Actogram	The graphical display of a time series along two time axes. The duration of a cycle (or predicted duration of a cycle) determines the length of each plot row (x axis); successive cycles are plotted on successive lines (along the y axis).
Entrainment	The synchronisation of a circadian oscillator to an external time-giving cue (see 'Zeitgeber' below). During a steady entrainment regime, the period of the entrained clock conforms to that of the external cue to establish a stable phase relationship.
Free run	The state of a self-sustaining oscillation in the absence of Zeitgebers.
Masking	A disruption in the expression of overt rhythms caused by an external agent without any direct effect on the circadian pacemaker. In the context of locomotor behaviour, masking is typically observed at points of rapid environmental transition, which acts to startle the animal.
Zeitgeber (ZG)	From the German, 'Time Giver'. A rhythmic environmental cue, which acts to regulate circadian rhythms, synchronising them to a 24 hour schedule.
Zeitgeber time (ZT)	A standard of time based on the period of a Zeitgeber. For diurnal organisms, Zeitgeber time zero (ZT 0) usually defines the ON phase of the entraining signal (e.g. onset of lights or warmth).

Table 1.2: Relevant circadian terminology

riously, the *P. andruzzii* clock is not temperature compensated, which likely reflects a lack of selective advantage for this feature at constant temperatures underground.

Similar principles are echoed in the arctic reindeer, *Rangifer tarandus*, which shows little-to-no clock mechanisms (Lu et al., 2010). It is believed that the lack of a strong circadian rhythm here arises from living in a part of the world that experiences long stretches of constant darkness and light throughout the year. Interestingly, *Rangifer tarandus* is acutely responsive to light-dark cycling during the equinoxes, however evidence suggests this is driven by a direct melatonin-mediated mechanism, rather than by true circadian clocks. (In this case, the stimulus-response argument of de Mairan's 18^{th} century opponents would have been correct).

Today, chronobiology has permeated many branches of the life sciences, pro-

viding context and mechanistic insight to countless areas, including sleep (Paceschott and Hobson, 2002), metabolism (Bass, 2012; Eckel-Mahan and Sassone-Corsi, 2013), ageing (Froy and Miskin, 2007; Tevy et al., 2013), development (Dekens et al., 2003; Kyriacou et al., 1990), tissue regeneration (Matsuo, 2003; Plikus et al., 2015), immune response (Keller et al., 2009; Wang et al., 2011), and cognition (Kyriacou and Hastings, 2010; Schmidt et al., 2007). This diverse biological influence is also highlighted by the 24-hour rhythmic activity of a number of genes: around 30-64% in cyanobacteria (Ito et al., 2009; Vijayan et al., 2009), 20-25% in Neurospora (Correa et al., 2003; Dong et al., 2008), 6-15% in Aradopsis (Covington and Harmer, 2007; Harmer et al., 2000), 5-25% in Mouse (Hughes et al., 2009; Koike et al., 2012; Panda et al., 2002; Storch et al., 2002; Vollmers et al., 2012), and 2% in the brain of *Drosophila melanogaster* (Hughes et al., 2012; Rodriguez et al., 2013). These proportions are in fact likely to be under-representative of total molecular rhythmicity due to the specificity of tissues analysed, and the existence of post-transcriptional regulation.

Of particular interest to clinical research, and funding bodies alike, is the ever-present signature of circadian dysfunction in a host of human pathologies, from cancer (Kelleher et al., 2014), metabolic syndrome (Maury et al., 2014), and cardiovascular disease (Morris et al., 2016), to psychosis and disorders of mood and sleep (Peirson and Foster, 2015).

The near ubiquitous presence of circadian systems in nature provides not only an indication of their selective benefit, but also an extensive range of model organisms for experimental research. Attractive from a biological point of view is that the shared evolutionary history of these systems makes investigation of one species likely to yield valuable insight into many others, including, of course, our own human functioning. With this luxury in mind, I have chosen to exploit the well-established *Drosophila melanogaster* as a versatile model organism that lends itself both to experimental and theoretical examination.

1.2 The *Drosophila* Circadian System

Genetic malleability and a readiness for imaging has made the fruit fly an essential model organism in the study of circadian timekeeping. Short generation time and a fast rate of reproduction promote high-throughput data collection, while behaviours exhibit strong circadian characteristics. Clock networks in *Drosophila* also pose a suitable balance between simple and complex, presenting a tractable system that promises to yield insight into more general circadian principles. In this section, I will discuss the architecture of this biological model.

1.2.1 Overview

The *Drosophila* circadian clock consists of central and peripheral cellular pacemakers, which coordinate to produce 24-hour rhythms in physiology and behaviour (Fig. 1.1). While the capacity to keep time is endowed at the level of individual cells, coordinated circadian behaviour in *Drosophila melanogaster* emerges from the concerted activity of a network of ~150 clock neurons located in the fly brain. This network is orders of magnitude smaller than the central clock in mammals (~20,000 neurons in size; Mohawk et al. 2012), yet there exists an astonishing level of functional and anatomical similarity between the phyla (Fig. 1.2; reviewed by Helfrich-Förster 2004).

The output of the central clock drives circadian behaviours, such as the daily rhythms of locomotor activity, which forms a key focus in this thesis (Fig. 1.1). A classical view of this central clock network describes a strict hierarchical organisation, with oscillatory 'masters' imposing rhythmicity on a more passive remainder of the network (Renn et al., 1999). However, recent work reveals greater autonomy within the brain clock, as independent neuronal subgroups are each capable of driving locomotor rhythms (Yao and Shafer, 2014).

Outside the brain, the existence of peripheral clocks in tissues around the body appear important for local temporal governance of biology (though less in known about the role of this wider peripheral network).

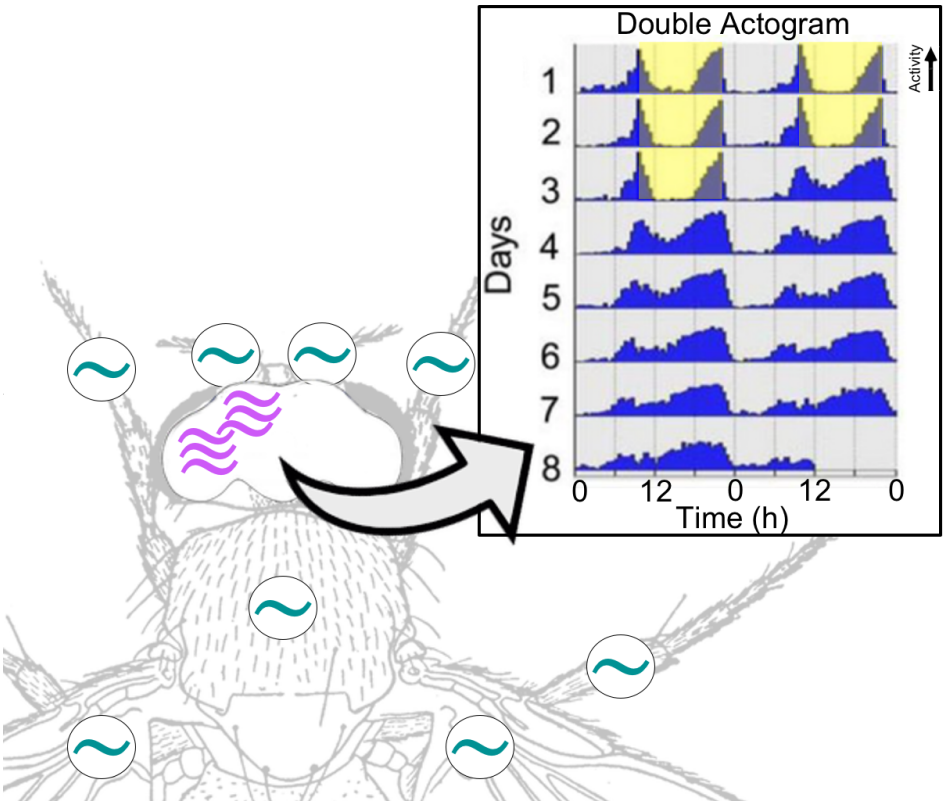

Figure 1.1: Schematic overview of the *Drosophila* circadian system. A central clock in the fly brain, comprising multiple independent oscillatory subunits (pink; left hemisphere only shown) drives daily rhythms of locomotor activity (dark blue trace). We can view these locomotor rhythms as a 'double actogram' which plots activity in time across a 48 hr period, such that the second half of each row is repeated in the first half of the row beneath. This allows the uninterrupted inspection of any 24 hr window along the length of a row. The central clock synchronises to the periodic variations of so-called Zeitgebers (see Table 1.2). This in turn synchronises locomotor rhythms. In the example data shown, wild type flies are exposed to cycles of light (yellow) and dark (grey); the central clock then continues to oscillate and drive rhythms of behaviour in the absence of external input. Beyond the central clock in the fly brain, multiple peripheral oscillators can be found around the body (green). Comparatively less is known about these peripheral clocks.

1.2.2 The Molecular Clock

To understand how the wider circadian system keeps time, we must first begin with the oscillatory building blocks that comprise it. Indeed, time is computed at the level of individual cells. Integral to these cellular oscillators is a molecular clock that is driven by the 24-hour rhythms of so called 'clock genes'. The first of these to be identified in *Drosophila* (and indeed the first of any organism) was the *period* gene (*per*).

At the final stage of metamorphosis in the *Drosophila* life cycle, the adult fly emerges from the hardened puparium in a process called eclosion. The process of eclosion is well-timed, coinciding with the moisture of the early morning (hence the Greek etymology of *Drosophila*, 'dew-loving'). In 1971, Seymour Benzer and his graduate student Ronald Konopka coupled an eclosion rhythm assay with a mutagenesis screen to identify three circadian mutants, all later shown to have targeted the *per* gene (Konopka and Benzer, 1971). They then went on to demonstrate defects in the timing of locomotor rhythms in these mutant flies.

Since Benzer and Konopka's seminal work, several other genes have been identified as key to circadian rhythmicity in *Drosophila*: these are *timeless* (*tim*) (Sehgal et al., 1994), *Clock* (*Clk*) (Allada et al., 1998), *cycle* (*cyc*) (Rutila et al., 1998), *doubletime* (*dbt*) (Price et al., 1998), *shaggy* (*sgg*) (Martinek et al., 2001), and the circadian photoreceptor, *cryptochrome* (*cry*) (Stanewsky et al., 1998). Together, these clock genes form the basis of transcription-translation feedback loops (TTFLs), in which gene protein products autoregulate their own transcription (illustrated in Fig. 1.2).

1.2.2.1 Transcription-Translation Feedback Loop

To illustrate the dynamics of the core TTFL in *Drosophila*, let us first consider the free-running case with no external sensory input. The basic helix-loop-helix-Per-Arnt-Sim (bHLH-PAS) transcription factors, CLK and CYC, activate transcription of the core clock genes, *per* and *tim*, through binding to E-box regions in the gene promoters (Allada et al., 1998; Darlington, 1998; Rutila et al., 1998). Subsequent translation of these mRNAs generates the protein products PER and TIM. The PER

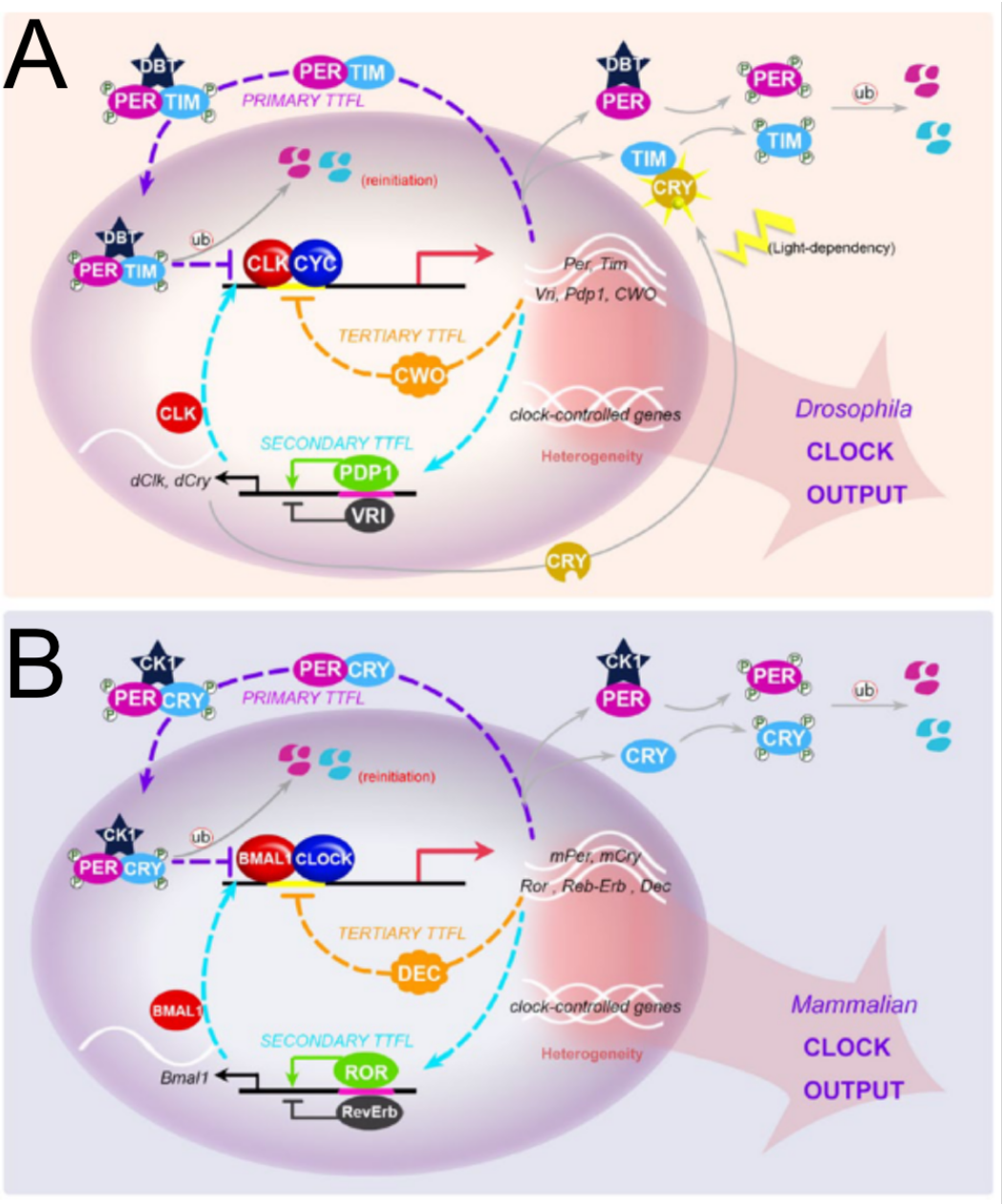

Figure 1.2: Comparison of the TTFL in central clock neurons of fruit flies and mammals. Molecular components are coloured according to functionally equivalent roles between species. Note the striking similarity between both systems, each of which include the presence of two interconnected negative and positive feedback loops. Note also the differing roles of CRY between fruit flies and mammals. Adapted from Ki et al. (2015).

protein is inherently unstable, and is quickly targeted for proteasomal degradation by the kinase DBT in a phosphorylation-dependent mechanism (Kloss et al., 2001; Price et al., 1998). However, through dimerisation with TIM, the PER-DBT complex is stabilised, leading to an increase in cytoplasmic DBT-PER-TIM (around 6-8 hr after the initiation of *per* and *tim* transcription) (Curtin et al., 1995; Zeng et al., 1996).

The DBT-PER-TIM complexes undergo nuclear localisation via the phosphorylation of PER and TIM by CK2 and SGG respectively (Akten et al., 2003; Lin et al., 2002; Martinek et al., 2001). Once in the nucleus, the complex binds CLK, disrupting CLK-CYC, and inhibiting transcription of the *per* and *tim* clock genes (Menet et al., 2010). PER and TIM protein levels consequently decline, reducing the nuclear abundance of DBT-PER-TIM, and ultimately removing the inhibition on CLK-CYC activity. The molecular cycle then starts again, *circa* 24 hours later.

1.2.2.2 Sensory Entrainment of the TTFL

The core TTFL is robust and persists in the absence of any external cues. However, according to the criteria outlined in Table 1.1 on page 18, the molecular clock must be entrainable to periodic environmental signals, or 'Zeitgebers' (ZGs). (It is this marriage of rigid and flexible traits that underscores a key selective advantage of timekeeping systems). Traditionally, research into the molecular workings of circadian clocks has focussed on photic input, revealing that light interacts with the core TTFL directly through activation of the circadian photoreceptor, CRY (Stanewsky et al., 1998). A conformational change in the light-activated CRY protein causes it to bind TIM, targeting it for ubiquitination via the F-box protein, JET (Koh, 2006; Peschel et al., 2009). This leads to rapid proteosome-mediated degradation of the TIM protein.

More recent work suggests similar principles at play during thermal input to the molecular clock (Tataroglu et al., 2015). In this study, evidence was provided to show that increased temperature leads to a rise in cytosolic calcium. A conformational change in Calmodulin (CaM) then causes it to bind TIM, which initiates a degradation process, this time mediated by the calpain protease, small optic lobes

(SOL). These results remain to be rigorously tested, but would identify the core clock protein, TIM, as a key molecular site of sensory integration between environmental light and temperature signals.

By adjusting the concentrations of core clock components in the TTFL, external cues can impart force to the endogenous rhythms of circadian clocks. In this way, such cues may act to synchronise, or 'entrain', clock systems – hence the etymology 'Zeitgeber', from the German, 'Time Giver'. How this temperature-sensitivity of the clock relates to the phenomenon of temperature compensation remains a topic of debate.

1.2.2.3 Additional Regulation of the Molecular Clock

This section has focussed on the core TTFL as a key mechanism underlying molecular rhythms in *Drosophila*. These principles are true across body tissues and phyla (see Fig. 1.2), and the TTFL is widely considered a universal building block of circadian clocks. For our purposes, and within the scope of this thesis, our discussion of molecular oscillatory mechanisms need go no further. However, it is important to note that the core TTFL does not represent a complete account of cellular timekeeping in *Drosophila*.

In order for the clock to truly be 'circadian', the molecular events discussed previously must take ~24 hours to complete. However, the kinetics of transcription, translation, and sub-cellular trafficking outlined above do not by themselves account for this 24-hour time frame. Further post-translational modifications are thus required to appropriately time the various components of the TTFL. We have already seen how kinases act to phosphorylate PER and TIM, leading to processes of degradation and/or nuclear localisation. Further ubiquitination of TIM is also essential for its light-dependent degradation, and in general, post-translational modifications play important roles in the precise timings of the molecular clock (for review, see Weber et al. 2011).

Additional to the core TTFL, there also exist parallel interlocked feedback loops (Fig. 1.2). For example, separate to its role in *per* and *tim* transcription, the CLK-CYC heterodimer also binds E-boxes in *vrille* (*vri*) and PAR domain pro-

tein 1 ε and δ (*Pdp1*ε/δ) (Cyran et al., 2003). The VRI protein goes on to bind VRI/PDP1-boxes in the *Clk* promoter, negatively feeding back to repress its transcription (Cyran et al., 2003; Glossop et al., 2003), whereas the PDP1ε/δ activates *Clk* transcription, forming a positive feedback loop component (Cyran et al., 2003). Finally, evidence suggests the existence of a third feedback loop in *Drosophila* involving the gene, *clockwork orange* (Kadener et al., 2007).

Examples of parallel feedback loops are present in the molecular clocks of many species and are believed to reinforce circadian oscillations, reducing dampening effects and creating resilience to external noise (Brown et al., 2012).

1.2.3 The Central Clock

Cytological staining for clock gene products has revealed the location of distinct neuronal subgroups within the central circadian network of the fruit fly. Each hemisphere contains 4 small ventral lateral neurons (s-LN_vs) expressing pigment dispersing factor (PDF), another small PDF-negative ventral lateral neuron referred to as the 5^{th} s-LN_v, 4 large PDF-expressing ventral lateral neurons (l-LN_vs), 6 dorsal lateral neurons (LN_ds), 17 group-1 dorsal neurons (DN1s), 2 group-2 dorsal neurons (DN2s), ~40 group-3 dorsal neurons (DN3s), and 3 lateral posterior neurons (LPNs). (See Fig. 1.3).

Despite the ability to identify distinct cell groups within the *Drosophila* central clock, the neurochemistry of this circuit is remarkably heterogenous. Signalling is achieved predominantly through the action of four neurotransmitters: glutamate (Hamasaka et al., 2007), neuropeptide F (NPF) (Lee et al., 2006), neuropeptide precursor-like protein 1 (Shafer et al., 2006), and the aforementioned pigment dispersing factor (PDF) (Helfrich-Forster, 1995; Park and Hall, 1998), which appears to mediate synchrony between different neuronal subgroups (Yao and Shafer, 2014).

While the precise functional connectivity within the network remains unclear, electrical coupling between neurons does appear to be important to normal functioning, as electrical silencing of neurons within the network impairs downstream behavioural rhythms (Nitabach et al., 2002).

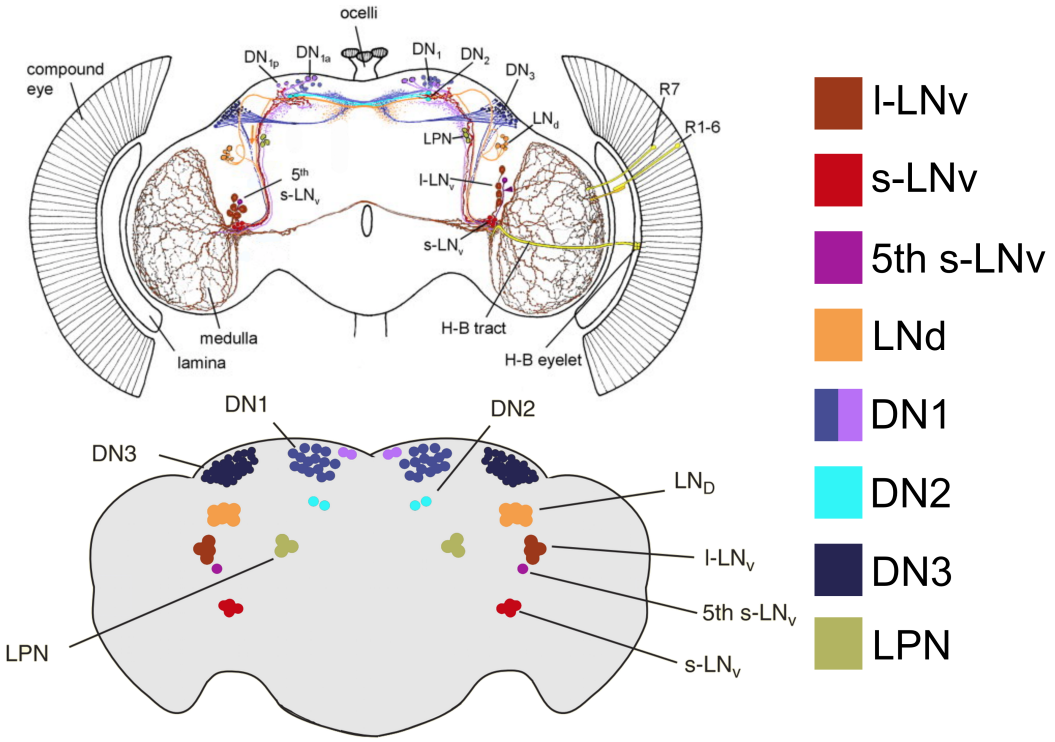

Figure 1.3: Architecture of the central clock network in *Drosophila melanogaster*. Adapted from Helfrich-Förster et al. (2007) (top) and Nitabach and Taghert (2008) (bottom). Neurons are coloured by subgroup (top, bottom) and connectivity between groups is shown (top).

1.2.4 Peripheral Clocks

Along with neurons in the central clock network of the fly brain, the circadian system includes peripheral clock networks in tissues around the body, including the retina, antenna, proboscis, leg, wing, malpighian tubules, gut, cuticle and reproductive organs (Giebultowicz and Hege, 1997; Giebultowicz et al., 2001; Ito et al., 2008; Plautz, 1997). These peripheral clocks are diverse, displaying varying levels of autonomy and involving different forms of molecular machinery (for review, see Ito and Tomioka 2016).

One striking difference between central and peripheral clocks concerns the role of CRY. In many peripheral clocks, CRY functions as both a circadian photoreceptor (akin to its role in the central clock) and a core component of the molecular machinery in the TTFL (see Fig. 1.4). In this latter role, peripheral CRY resembles more closely that which is observed in mammalian systems, where CRY plays a role

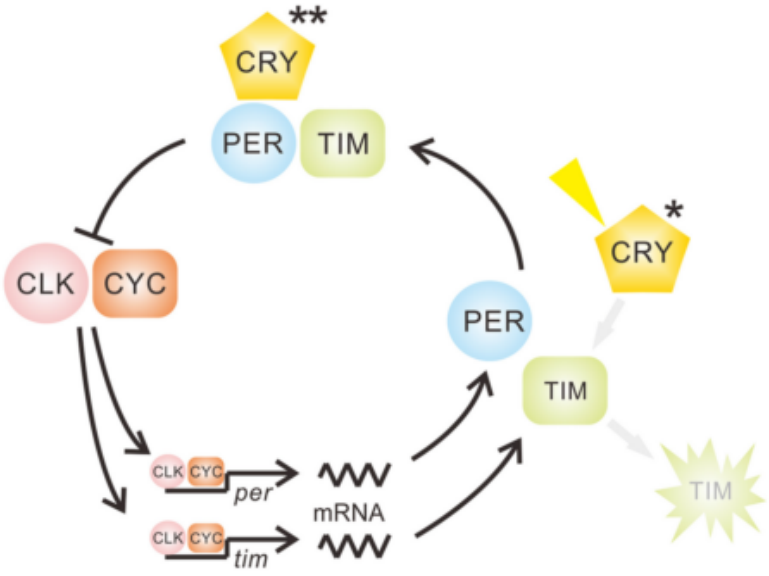

Figure 1.4: TTFL in the peripheral clock neurons of fruit flies. A diagram showing the molecular components of peripheral clocks in *Drosophila melanogaster*. Note the dual role or CRY as either a photoreceptor involved in the light-dependent degradation of TIM (*), or a core clock component analogous to that seen in the mammalian clock (**). Adapted from Ito and Tomioka (2016).

reminiscent of *Drosophila* TIM in the stabilisation of PER (see Fig. 1.2) (Okamura, 1999; van der Horst et al., 1999).

1.3 Locomotor Activity Rhythms

Circadian rhythmicity exists at the cellular and network level, yet it is appropriately-timed behaviours that make this system adaptive for the organism. The standard circadian behavioural readout in *Drosophila*, and indeed many other laboratory organisms, is locomotor activity.

1.3.1 Free running behaviour

The circadian tale emphasises endogenous rhythmicity as a key trait of clock-driven behaviours. Locomotor activity is no different. Following entrainment to a given environmental regime, *Drosophila melanogaster* continue to show robust rhythms of activity in constant conditions (Konopka and Benzer 1971; Konopka et al. 1989; see Fig. 1.1 for example). However, not all regimes are created equal. In constant light, flies quickly become arrhythmic, compared to more persistent rhythmicity in

constant darkness (Konopka et al., 1989). It should also be noted that a gradual decrease in activity levels is observed over time (this can be seen in any fly locomotor experiment conducted in constant conditions, and is shown in this thesis).

1.3.2 Photic Entrainment

The exact activity profiles displayed by *Drosophila melanogaster* depend on the precise environmental conditions and the genetic background of the fly (see Schlichting and Helfrich-Foerster 2015). One easy-to-implement regime – and consequently one of the most well-studied – implements a rectangular light-dark cycle, where lights are either on or off at a constant temperature. This light-dark (LD) regime is adequate to entrain the activity rhythms of *Drosophila melanogaster*. Indeed, Konopka and Benzer made use of these conditions in their seminal work (Konopka and Benzer, 1971).

Under photic cycles comprising 12 hr light and 12 hr darkness (12:12 LD), *Drosophila* display a characteristic bimodal activity pattern with a morning (M) peak and an evening (E) peak (See Fig. 1.5 A). One key feature of this locomotor profile is a characteristic ramping increase of activity preceding the environmental transitions (lights on/off). This feature of the activity rhythm highlights the action of an internal oscillator, rather than a simple stimulus-response relationship - the behaviour *anticipates* changes in the environment, rather than simply *reacting* to them. Such identifiably circadian components of the activity trace are useful for chronobiologists as light can also affect locomotor activity in a clock-independent manner (commonly referred to as 'masking'). For example, sharp fluctuations of light intensity can startle the flies, leading to spikes in activity levels. Thus, while it is clear that fruit fly locomotor activity is under strong circadian control, making it a suitable behavioural readout of clock function, we must take into account these non-circadian effects when interpreting the data.

1.3.3 Thermal Entrainment

Just as the molecular clock has been shown capable of entraining to both photic and thermal signals, so too do the rhythms of locomotor activity in the fruit fly

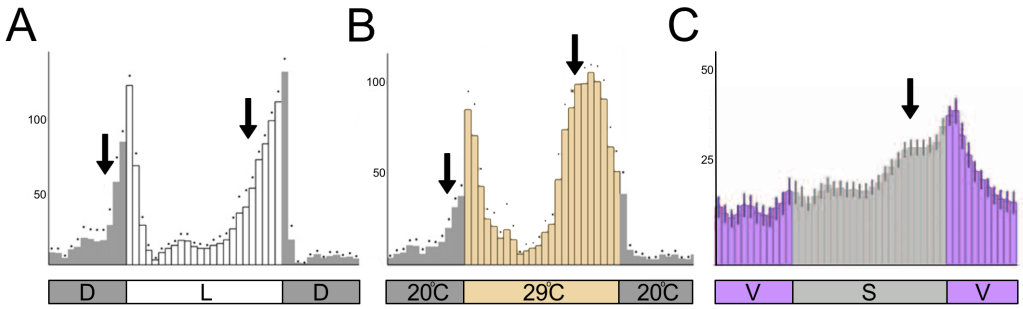

Figure 1.5: Circadian locomotor profiles during entrainment to light (A), temperature (B), and mechanosensory stimuli (C). (A) Entrained activity of wild type flies (w^{1118}, n = 16) to 12:12 LD at 25°C (adapted from Zhang et al. 2010). (B) Entrained activity of wild type flies (w^{1118}, n = 16) to 12:12 TC (20°C:29°C) in constant darkness (adapted from Zhang et al. 2010). (C) Entrained activity of wild type flies (Canton S, n = 75) to 12:12 VS in constant darkness at 25°C (adapted from Simoni et al. 2014).

(Glaser and Stanewsky, 2005; Sehadova et al., 2009; Wheeler et al., 1993). Again, the precise shape of the locomotor trace depends on the specific temperature regime used. For example, the evening peak of activity appears advanced during temperature cycles around a lower mean versus those around a higher mean (Gentile et al., 2013).

Furthermore, there is compelling evidence that a gradual, more naturalistic, transition between warm and cold temperatures leads to the emergence of a distinct activity peak in the afternoon (Fig. 1.6; Green et al. 2015; Vanin et al. 2012). In most experimental temperature regimes, however, a bimodal activity pattern is commonly observed, where anticipatory increases in activity can be seen preceding morning and evening (beginning/end of warm phase; Fig. 1.5 B).

Clearly, the study of circadian behaviour under more artificial laboratory conditions does not provide a complete view of the fruit fly clock. As ever, the challenge is in identifying those environmental conditions that adequately probe the phenomena we wish to study, while minimising complexity in order to aid comprehension. Furthermore, these naturalistic field studies raise interesting questions over phase relationships between photic and thermal signals. Fig. 1.6 shows an intrinsic degree of misalignment between light and temperature cycles (peaking at different times); it remains unclear how this is interpreted by the clock system.

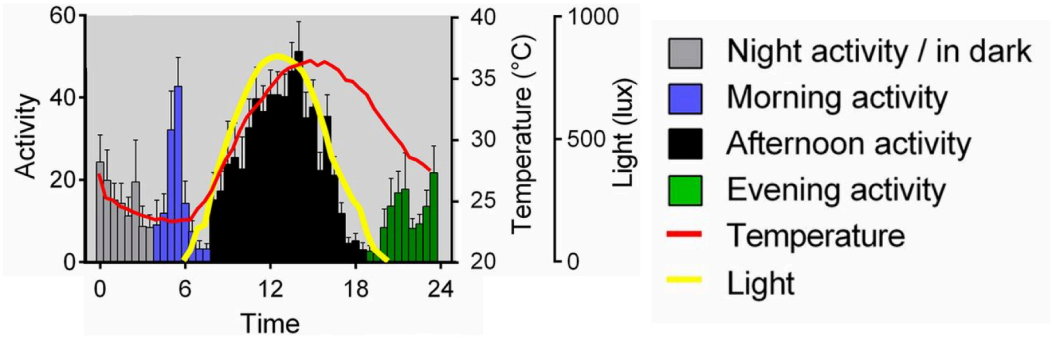

Figure 1.6: *Drosophila* circadian locomotor behaviour during natural conditions. Activity of male wild type flies (n = 9) collected as natural isolates from Houten, The Netherlands, recorded in natural conditions on 28th July, 2007, in Treviso, Italy. Note the presence of an afternoon peak of activity, distinct from the morning and evening peaks. Adapted from Green et al. (2015).

1.3.4 Mechanical Entrainment

While circadian entrainment to light and temperature are by far the most well-studied, the fruit fly circadian clock appears to adjust its phase to other stimuli, including social, olfactory, and mechanical cues (Fujii et al., 2007; Levine et al., 2002c; Simoni et al., 2014). An in-depth discussion of these novel ZGs lies beyond the scope of this thesis. However, the presence of a mechanosensory input to the fly clock is of particular interest.

In a recent study, mechanosensory rhythms comprising 12 hr of vibrations followed by 12 hr of silence (12:12 VS) were shown to be capable of entraining both locomotor activity and PER rhythms in wild type *Drosophila melanogaster* (Fig. 1.5 C; Simoni et al. 2014). Curiously, entrainment to these mechanosensory cycles required in-tact proprioceptive chordotonal organs, as genetic ablation of these structures abolished entrainment. Importantly, we can once again see an anticipatory increase of activity preceding the transition from silent to vibrational phase, revealing a clear circadian component to the locomotor waveform.

1.4 Multisensory Entrainment

We have seen that both at the level of molecular and behavioural oscillations, the circadian clock in fruit flies entrains to sensory cues in the environment. Research in the field has traditionally focussed on modulating just one of these cues, and

holding others constant. While this approach is intuitive, and has been successful at uncovering key mechanisms of sensory entrainment, it remains limited. In nature, sensory cues tend to fluctuate in tandem with one another. For example, as the Sun illuminates a habitat, it will also begin raising the surrounding temperature. Thus, circadian clocks have evolved in, and are subject to, multisensory challenges.

1.4.1 Experimental Findings

Multisensory treatment of circadian clocks has begun to emerge in the last decade. Yet such a view remains only to have been embraced by only a relatively small number of studies (Currie et al., 2009; Miyasako et al., 2007; Yoshii et al., 2010). This work represents a firm foundation for the experimental investigations presented in this thesis.

As we saw previously, one avenue to investigate multisensory integration in the fly clock might be to assess this system during naturalistic ZG phase relationships (Green et al., 2015; Vanin et al., 2012). Alternatively, a sensible approach to explore any neurobiological system is to ask the question, 'what happens when there is disagreement between inputs?' In the circadian setting, this question manifests itself in the form of deliberate misalignments between two entraining signals. Accordingly, research on this topic has implemented environmental regimes that present out-of-phase light and temperature cycles (focussing on these two sensory modalities due to their previously-discussed prominence in the circadian literature).

1.4.1.1 Molecular Rhythms

Immunohistochemistry has revealed differential entrainment of individual cells in the clock network, with a light preference displayed by CRY-containing neurons, including the l-LN_vs, s-LN_vs, 5th s-LN_v, and the CRY^+ DN1s, as well as the more heterogeneous CRY^+ and CRY^- LN_d and DN3 subgroups (Fig. 1.7; Yoshii et al. 2010). In contrast, temperature preference was seen in neurons lacking the CRY photoreceptor, including the LPNs, DN2s, and the CRY^- DN1p cells (Fig. 1.7; Yoshii et al. 2010). These results are consistent with those obtained by Miyasako et al. (2007), and go some way toward describing a network that is sub-specialised

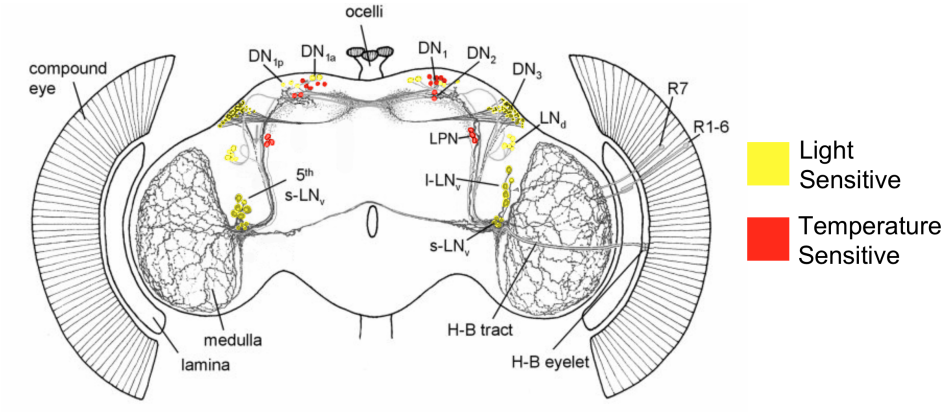

Figure 1.7: Sensory preferences of neuronal subgroups in the central clock of *Drosophila melanogaster*. A schematic representation of the findings from Yoshii et al. (2010) and Miyasako et al. (2007), showing light preference (yellow) and temperature preference (red) of different neuronal subgroups during misaligned LD:TC. Adapted from Helfrich-Förster et al. (2007).

to process different sensory signals (Fig. 1.7). Furthermore, this highlights the central clock as a network of oscillatory subunits, many of which are independently capable of orchestrating locomotor activity (Yao and Shafer, 2014).

While a small number of studies have begun to investigate the behavioural and molecular response of the central clock network to conflicting sensory input, the response of peripheral clocks remains entirely unstudied.

1.4.1.2 Behavioural Rhythms

While molecular investigation hints at the complex processing of conflicting input signals by specialised neuronal subgroups within the clock network, behavioural findings paint a simpler picture. Indeed, results to date suggest *Drosophila* locomotor rhythms entrain preferentially to photic signals.

Light dominance in the circadian system of fruit flies is a traditionally popular view in the field. During both 12 hr (Yoshii et al., 2010) and 6 hr misalignments of LD:TC (Miyasako et al., 2007), locomotor behaviour in wild type flies appeared to entrain preferentially to the light cue. However, in another study investigating 12 hr misaligned LD:TC, temperature was found to have a more substantial circadian effect, advancing the onset of evening locomotor activity (Currie et al., 2009). This is perhaps more consistent with the previously-discussed field studies exploiting natu-

ralistic environmental fluctuations (Fig. 1.6), which also suggest a more prominent role of temperature in locomotor entrainment (Vanin et al., 2012). The situation thus remains unclear.

1.4.2 Theoretical Considerations

It has been argued that any information processing system can be described and understood at three distinct levels (Marr, 1982). These are defined as computational (what problem does the system solve), algorithmic (by what rules and methods does the system operate), and implementational (how is the system physically realised). Having explored the implementational aspects of the circadian clock in *Drosophila melanogaster*, let us briefly address the remaining two components of David Marr's tri-level hypothesis of understanding.

A key role of the nervous system is to estimate the state of the world from noisy or incomplete data. The nature of this state can be separated into observable variables (things we perceive) and hidden variables (things we must infer from observable variables). A common illustration of this idea is to imagine a cat − it may meow (the observable variable), but we cannot directly sense that it is hungry (unobservable variable), only infer this state from its behaviour.

The same principle can be applied to the circadian clock. ZGs act as observable variables, from which the organism must compute the otherwise unobservable time of day. However, as we have seen in this chapter, the *Drosophila* clock can process different sensory modalities. We may therefore define the computational role of the system as follows: to sense time from various, potentially conflicting, sources of information, and use this to drive appropriately-timed output.

Bayesian integration provides an optimal algorithmic method by which to combine different sources of information. Here, the goal of a Bayesian observer is to compute the conditional density function, specifying the probability of the time of day given information from light and temperature:

$$p(t|L,T) = \frac{p(L,T|t)p(t)}{p(L,T)} \tag{1.1}$$

where L and T denote the light and temperature signals respectively, and t is the time of day. The likelihood function $p(L,T|t)$ describes the likelihood of sensing a particular light and temperature signal for a given time of day. $p(t)$ is the prior probability of the time of day, and $p(L,T)$ is a normalising constant.

In the case that variations in the light and temperature signals are independent given the true time of day, we can decompose the likelihood function in Eq. 1.1 into the product of the likelihood functions associated with each cue:

$$p(L,T|t) = p(L|t)p(T|t) \tag{1.2}$$

This is perhaps an over-simplification. Indeed, variations of light and temperature are likely to be co-dependent. Moreover, past information on the ZGs will likely enter into the final computation of time in the form of a Bayesian prior. I will return to these ideas in Chapter 8. However, for now, the concepts presented in their current form provide a useful conceptual framework for subsequent experimental investigations.

In the case where there is no prior information (i.e. $p(t)$ in equation 1.1 is uniform), an optimal integrator is one that calculates the time of day from the mode of the likelihood function in Eq. 1.2. As light is traditionally considered a more potent entraining stimulus than temperature (see page 33) we might imagine that this is the more reliable of the two cues, which would be reflected by a smaller variance in the probability distribution, $p(L|t)$; conversely, $p(T|t)$ would follow a broader distribution. An optimal Bayesian estimation of the time of day is therefore some combination of the time specified by light and temperature, weighted according to the reliability of each cue. Studies of multisensory processing in circadian clock systems would benefit from keeping these ideas in mind.

Chapter 2

Background: Computational Chronobiology

As we have seen, the biology of circadian clocks is complex, and requires understanding at many different levels. This complexity makes it difficult to theorise over system mechanisms and functions. As part of a more integrated approach, we can represent our beliefs as a mathematical model, describing the system as a set of equations. The variables of the equations may represent specific neurobiological quantities (e.g. the concentration of a given protein) or they may lie in a more abstract space. Likewise, the equations themselves denote how different quantities interact with one another according to the theories being expressed by the model. In this way, mathematical modelling provides biology with a vital framework to explain and predict.

A challenge in the construction of mathematical models is abstracting a system we seek to describe into something that is analytically or computationally tractable. With the advent of more powerful computing, these limitations have reduced somewhat, allowing us to build models of increasing size and complexity. Yet, for our efforts to possess relevance to the study of biology, model architectures must always be rooted in experimentally-derived knowledge. In this way, there exists an intrinsic inter-dependency: a stepwise process of experimentation guiding theory, and theory providing insight and predictions to direct further experimentation. This reasoning has formed a core element of my academic outlook and attitude to research

throughout my PhD.

In this chapter, I wish to provide an overview of the types of numerical approaches employed in the field of chronobiology. My aim is to provide accessible information to those non-experts wishing to learn more about the area, while also establishing a background on which to base my own contributions (see Chapter 8). I begin with a discussion of the current tools for quantitative analysis of circadian data, focussing on techniques relevant to my research question, and drawing attention to their strengths and weaknesses. I then discuss individual oscillator models of varying complexity, before developing these ideas into network-level phase models of the clock. Finally, I consider phase response dynamics of oscillatory systems.

In the spirit of illustrating these mathematical approaches in a more tangible and affable manner, many have been implemented as explanatory examples. In this way, the reader should hopefully become more convinced of their immediate availability to the chronobiological investigator.

2.1 Phase Analysis of Circadian Data

What exactly do we mean by the term 'phase'? It is ambiguous. Indeed, there are a number of closely-related interpretations: 1) as a quantity by which state of motion is represented, 2) the argument of a sinusoidal function, or 3) to define the relationship between a periodic waveform and a fixed reference point in time. For most purposes within chronobiology, 'phase' is used to denote the fraction of a complete cycle along a periodic signal. This is a useful definition for many circadian treatments, and I will adopt it in this thesis. However, it pays to remember that the notion of phase is in fact only well-defined for trigonometric functions (such as sine or cosine waves).

Circadian clocks are often considered 'limit cycle oscillators' in that small perturbations are tolerated by the system, which always relaxes back to some constant amplitude and period. Yet, while amplitude and period of circadian rhythms display these steady-state tendencies, the phase of oscillations is acutely sensitive to perturbing forces. It is perhaps then no surprise that phase considerations of cir-

cadian clocks have dominated quantitative investigations in the field. Indeed, of particular relevance to this thesis is the idea that sensory cues exert entraining effects primarily though adjusting clock phase. Let us consider existing approaches for evaluating circadian phase in time series data.

2.1.1 Parametric Approaches

Given the clear relationship between phase and trigonometric functions, it is perhaps no surprise that a popular method for evaluating phase in circadian rhythms involves fitting a cosine function to the data (Halberg et al., 1967).

$$f(t) = M + A\cos(\frac{2\pi}{\tau}t + \varphi) \tag{2.1}$$

where t is time, M is a constant displacement of the function, also known as the midline estimating statistic of rhythm (MESOR), A is the amplitude, τ is the rhythm period, and φ is the phase of the peak of the fitted curve.

Least squares regression may be used to calculate the best fit parameters of the cosine function, and thus the phase of the biological rhythm may be inferred. This method breaks down slightly as the waveform deviates from a sinusoid (as we have seen already in the case *Drosophila* locomotor rhythms). In these cases, the chronobiologist may appeal to Joseph Fourier's revolutionary insight that any time series, regardless of shape, can be characterised by a series of sine and cosine waves of varying frequencies (Fourier, 1822). Indeed, more complicated harmonic oscillator models have been used to analyse phase of circadian datasets (Brown and Czeisler, 1992). Caution must be used, however, as gratuitous addition of harmonics can be seen as arbitrary (Wang et al., 2003), and does not represent any specific underlying biological process. Also, with an increase in the order of the harmonic function comes a reduction in our ability to infer phase from the model; it also becomes unclear how best to interpret circadian phase when these multiple frequency components can themselves be misaligned relative to one another.

Another important limitation of trigonometric fits is that they are stationary,

meaning that their parameters remain constant over time. This assumption does not hold for many circadian datasets, most notably in the case of entrainment, where the phase of the oscillation is adjusted by external signals. For this reason, linear fits of this type are generally restricted to small time frames.

One approach to handle the non-stationary aspects of circadian datasets is to use wavelet methods, which were originally designed to handle such time series (Percival and Walden, 2000). These methods have not yet been applied *en masse* in the field of chronobiology – perhaps due to inaccessible communication of the mathematics involved – but they have shown some early promise (Leise, 2013; Price et al., 2008).

2.1.2 Non-parametric approaches

The parametric methods described above require a number of assumptions to be made (e.g. the circadian waveform is well-described by some combination of harmonic oscillator functions). These assumptions may well hold true in many cases, but there is also a risk that researchers can force-fit their data, and in doing so draw erroneous conclusions.

Non-parametric methods go some way toward mitigating these issues at the expense of potentially reducing analytical power. It is therefore sometimes desirable to employ non-parametric approaches, and in doing so, let the data 'speak for itself'. One such application to the analysis of phase involves smoothing the time series data using a low-pass band filter (Levine et al., 2002a). An appropriately chosen bandwidth acts to extract the circadian components of the signal, from which some phase landmark (usually the peak) can be empirically measured.

A benefit of this non-parametric approach is that it makes no strong assumptions on the underlying period (beyond the choice of bandwidth) and therefore can accommodate non-stationarities in the data.

2.1.3 Current Challenges

Existing methods of phase analysis are imperfect. Parametric approaches suffer from the need for multiple assumptions, including specification of the underlying

functional form of the rhythms, and often that the time series must have stationary statistical properties. Non-parametric approaches can alleviate some of these problems, but they too come with disadvantages. For example, without a defined waveform, it is impossible to extract a more continuous measure of circadian phase from the data. The researcher is restricted to the use of signal landmarks (such as the peak), which greatly reduces temporal resolution and could stifle more in-depth analysis of sensory entrainment (e.g. circadian phase could advance and then delay between two peaks without detection).

Both parametric and non-parametric phase analysis methods are sensitive to the waveform of the signal. Let us imagine, for example, data traces that are all free running with the same phase and period, differing only in the precise shape of their bimodal waveform. (Fig 2.1 illustrates such a collection of data). Modelled on fly locomotor data of the kind shown in Fig. 1.5 (page 31), the trace in the top panels of Fig. 2.1 exhibits lower activity in the morning, and the trace in the bottom panels exhibits slightly greater activity levels in the morning. Simply put, the activity waveforms show variability between individuals. It is clear that the phase measured by the simple harmonic fit is advanced as the morning activity increases (Fig 2.1). This is because the fitting procedure will seek to minimise the distance between the cosine curve and the data. This makes comparison of phase measurements between animals with differing behavioural waveforms challenging – be they of the same or disparate genotypes.

The non-parametric approach fairs better in the example shown in Fig. 2.1, with the peak phase measurement consistent between traces. However, we might still question the arbitrariness of using this signal landmark. The meaning behind peak activity may not translate between genotypes and environmental conditions. It is also unclear how these methods accommodate periodic traces in which there are multiple prominent peaks per cycle, or indeed no discernible peak in the waveform, as has been shown to occur in *Drosophila* locomotor activity traces (Fig. 2.1; Harper et al. 2016; Vanin et al. 2012).

Finally, both the parametric and non-parametric methods outlined are suscep-

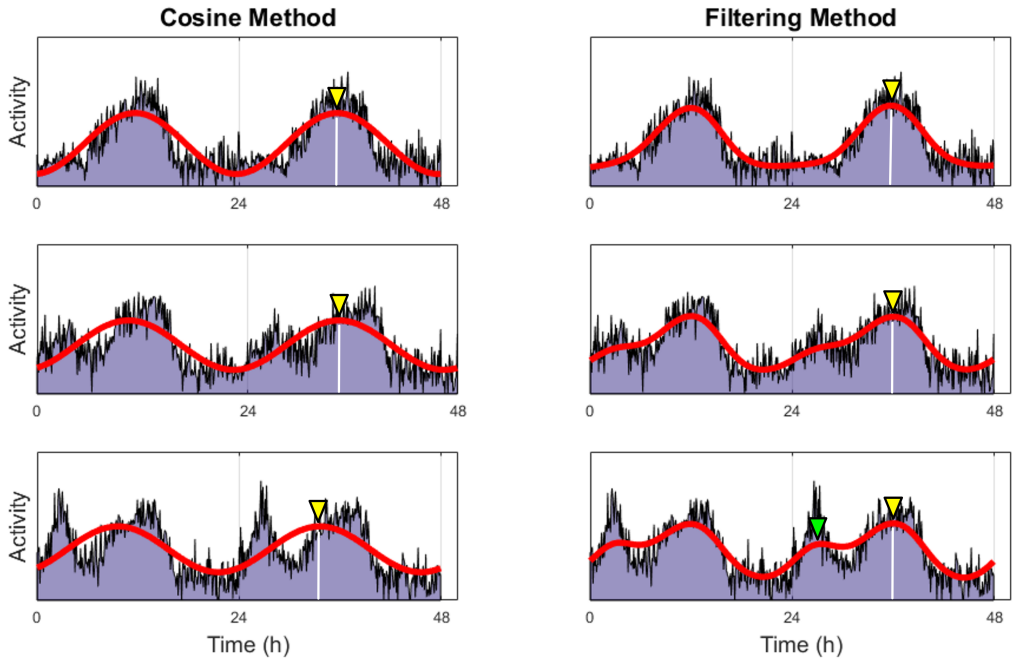

Figure 2.1: Waveform-sensitivity of existing phase analysis methods. The figure shows three example locomotor traces with the same phase and period, but increasing levels of morning activity components to their bimodal waveforms (top, middle, bottom). The example traces are analysed with parametric and non-parametric methods (left, right). Location of the phase measurement in all cases is shown by a white line and yellow triangle marker and restricted to the second day for ease of viewing. As the waveform of the locomotor behaviour changes to display greater levels of morning activity, peak phase of the simple cosine fit is advanced, whereas peak phase of the filtering procedure remains constant. However, with increased morning activity, the filtering procedure identifies a second peak (green triangle marker), which could act to confuse an automated protocol.

tible to noise in the data. Chronobiologists typically address this issue by averaging across populations, however this necessarily prevents phase analysis of individual animals.

Comment

It is important to stress that my emphasis on phase-abstraction does not negate the utility of other circadian metrics. Indeed, it would be foolish to make such an argument – amplitude, period, waveform, and rhythm strength are all key features of circadian clocks, and remain essential to wider chronobiological investigation. It is

undeniable, however, that considering circadian clocks in terms of their phase has proven to be an intuitive and useful way to analyse this system. Moreover, it is of special importance to studies of sensory entrainment and mathematical modelling.

2.2 Dynamical System Models

A dynamical system model is a mathematical description of the evolution of some value in time. In the context of circadian clocks, this value could be clock phase, or even the concentrations of molecular components in the cellular clock. Many attempts to model the circadian clock in *Drosophila* have been deterministic in nature, meaning that once parameter values and initial conditions are established, the system evolves in a unique and predictable manner. In this section, I will discuss key examples of these.

2.2.1 Gene-Regulatory Models

One of the earliest examples of mathematical modelling in chronobiology dates back to 1957, where a van der Pol oscillator model was used by Colin S. Pittendrigh − a man credited today as a father of chronobiology (Pittendrigh and Bruce, 1957). While the parameters of the van der Pol model did not directly correspond to molecular clock components, it was able to simulate circadian characteristics, such as free-running rhythms with a 24-hour period, and a capacity for entrainment to a periodic ZG.

A few years later, Brian C. Goodwin developed a more biologically relevant oscillator model based on a negative feedback loop comprising a chain of three variables, X, Y, and Z, where Z inhibits the production of X (Goodwin 1965; Fig. 2.2). This has since been widely used in the circadian field (Drescher et al., 1982; Ruoff et al., 1999). The fact that a negative feedback loop was proposed around 30 years prior to the discovery of the TTFL in *Drosophila* (Hardin et al. 1990; see Chapter 1) only highlights the strength of mathematical modelling in the process of biological investigation.

Over the years, deterministic models of this form have been improved and adapted according to new information obtained from experimental work (again il-

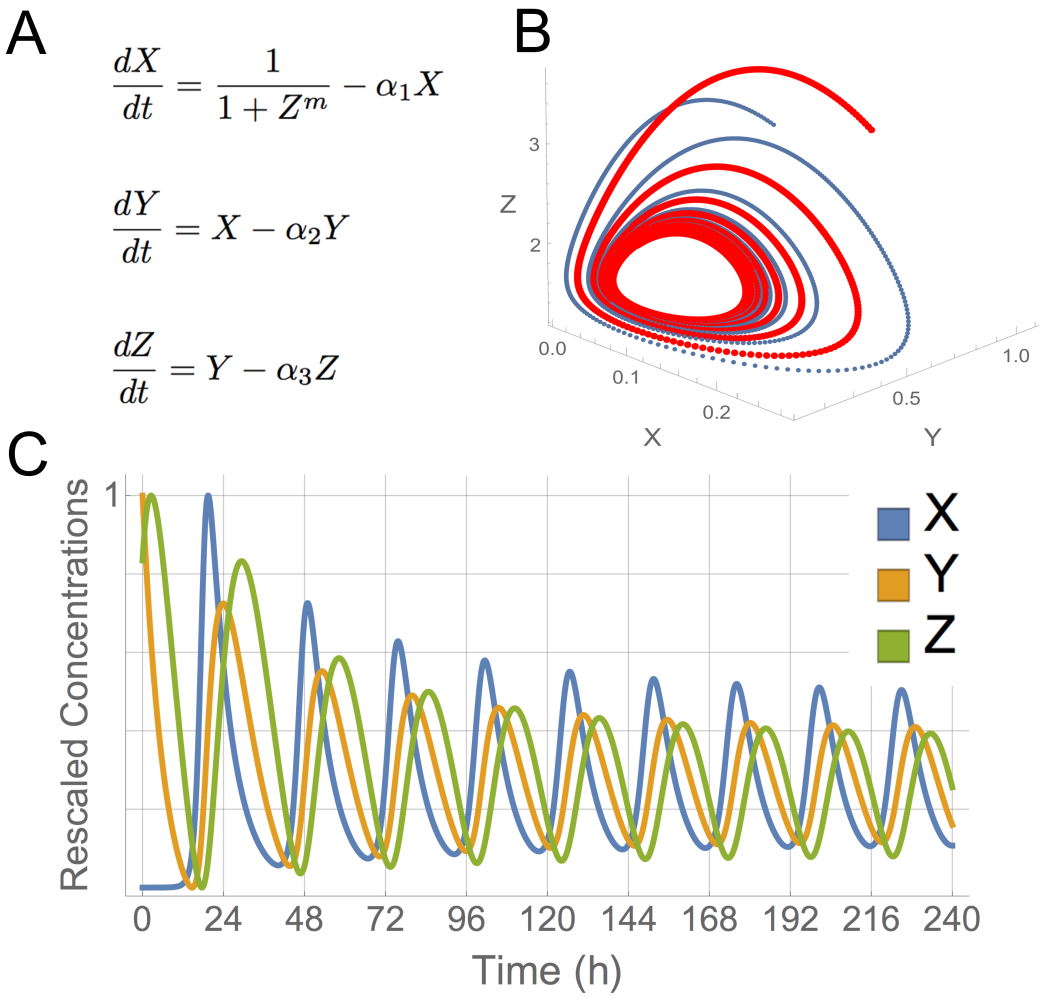

Figure 2.2: Goodwin oscillator simulations. A) Goodwin model describing the rate of change of three molecular species: X, Y, and Z. Decay rates represented by α and number of repressor molecules shown by m (see Drescher et al. 1982). B) Limit cycle simulation of Goodwin model, where $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$, and $m = 9$. System starts in one of two initial conditions (Blue: $X(0) = 0, Y(0) = 1, Z(0) = 3$; Red: $X(0) = 0.1, Y(0) = 1.1, Z(0) = 3$). C) Time series simulation of Goodwin model where parameters are the same as those used in (B). Molecular concentrations rescaled to fall between 0 and 1 for ease of visualisation. Note amplitude and period relax toward steady-state values. All simulations performed in *Mathematica*.

lustrating the reciprocity between experimentation and theory). These range in complexity from 5-variable (Goldbeter 1995; Fig. 2.3 left), 10-variable (Leloup and Goldbeter 1998; Fig. 2.3 middle) and 19-variable gene-regulatory models (Leloup and Goldbeter 2003; Fig. 2.3 right), and typically incorporate non-linearities in the form of Michaelis-Menten and Hill kinetics.

These in-depth biochemical models are certainly attractive in their mechanistic explanations of circadian dynamics. Indeed, a 5-variable model of PER oscillations in *Drosophila* was capable of sustained limit cycle oscillations, and predicted a dependence of free running period on PER degradation rate, providing a tentative explanation for discrepancies between different *per* mutants (Goldbeter, 1995; Konopka and Benzer, 1971). A later incarnation of this model included the role of TIM (see Chapter 1), with an associated light-dependent degradation parameter to permit photic entrainment. This model exhibited phase response dynamics that matched experimental findings in both wild type and *per* mutant flies (Leloup and Goldbeter 1998; Fig. 2.6 D).

Mechanistic descriptions of circadian systems can be useful, but their firm foundation on biological observations can also be a weakness. Continued experimentation leaves theoreticians little choice but to incorporate additional components to their models. At the time of writing, I am aware of a biochemical model of the mammalian clock that comprises 73 different components (Forger and Peskin, 2003). If we follow this trajectory toward the realms of *reductio ad absurdum*, we start to imagine mathematical systems that will eventually rival in complexity the very biological entity they seek to model. At which point, we might question the utility of such approaches for the tangible understanding of circadian principles. Perhaps it is time to explore new avenues (see Chapter 8). The sophistication of David Marr's computational level of explanation could provide the clarity we need (Section 1.4.2).

2.2.1.1 Adding Noise

Further complexity can be added to gene-regulatory models in the form of noise. Biological systems typically exhibit stochasticity, limiting the usefulness of purely

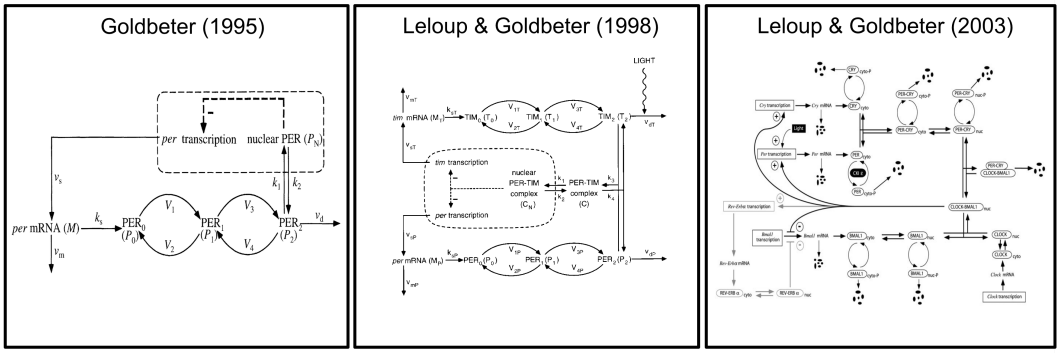

Figure 2.3: The growing complexity of gene-regulatory models. A 5-variable (left) and 10-variable (middle) model of the *Drosophila* circadian clock (adapted from Goldbeter 1995 and Leloup and Goldbeter (1998) respectively). The latter was updated to include the role of *timeless* in the molecular clock of the fruit fly (see Fig. 1.2 for clarification). A 19-variable model of the mammalian clock (right) shows the action of PER, CRY, BMAL1, CLOCK, and REV-ERBα, along with post-translational regulation of these proteins (adapted from Leloup and Goldbeter (2003))

deterministic models to describe and explain them. Gene expression is an inherently noisy process, and can vary between cells in an animal (Raj and van Oudenaarden, 2008). It should be no surprise then that circadian oscillations, which are based on transcriptional-translation mechanics, also display a degree of intrinsic noise. One option is to discount this noise at the level of reaction kinetics by appealing to the law of large numbers, where the importance of individual variations are negligible. However, this assumption may not hold for the potentially very low concentrations of clock gene mRNA and protein products in cells.

Accordingly, and in the spirit of grounding assumptions in biology, many of the models discussed in the previous section can be adapted through the addition of noise to the reaction kinetics. One popular method of simulating molecular noise was pioneered by Daniel T. Gillespie, in which possible reactions at discrete time steps are assigned a probability proportional to the abundance of reacting molecules. Monte Carlo sampling specifies which reactions occur according to their respective probabilities, as well as the duration until the next reaction. Subsequent reaction products are used to update the molecular abundances in the system and the corresponding probabilities at each time step (Gillespie, 1977).

2.2.2 Network Representations

As discussed in Chapter 1 (page 21), the circadian system of *Drosophila* and many other organisms is a vastly inter-connected *network* of autonomous oscillators. Modelling the dynamics of clock cell populations is therefore vital to a complete mechanistic description of this system. Moreover, it seems reasonable to suggest that network models may be required for investigations of network characteristics, such as inter-neuronal coupling strengths and synchrony conditions.

One approach to mathematically represent clock networks is to construct in-depth biochemical models of the individual oscillatory subunits (discussed in Section 2.2.1), and connect these together. As one might expect, *in silico* simulations of this type can quickly become infeasible as the number of oscillators in the network grows.

An alternative approach has been to implement simpler abstractions of the individual oscillatory units. Indeed, the original Goodwin oscillator has seen a renaissance, where it has been adapted to include a neurotransmitter term that permits network-wide coupling between the simpler three-species cells models (Gonze et al., 2005; Gu et al., 2009; Locke et al., 2008). Further attempts have been made to reduce the dimensionality of more complex gene-regulatory models while retaining biological meaning in the parameters of the model (Indic et al., 2006; Taylor et al., 2008) .

Alternatively, more phenomenological models may be used. Here, individual oscillatory components are represented by only a small number of descriptive features. Pioneered by Art Winfree (Winfree, 1967), and subsequently formalised mathematically by Yoshiki Kuramoto (Strogatz, 2000), weakly coupled oscillator models (WCOs) have proven particularly successful in modelling multicellular circadian systems (Gu et al., 2016; Indic et al., 2008; Liu et al., 1997). Originally motivated by the phenomenon of collective synchronisation, WCOs describe individual oscillators purely in terms of their phase. Assuming only weak coupling between neurons, and similar natural (free-running) frequencies, long-term dynamics are given by equations of the following form:

$$\frac{\partial \theta_i}{\partial t} = w_i + \sum_{j=1}^{N} K_{ij} \sin(\theta_j - \theta_i) \tag{2.2}$$

where θ_i is the phase of oscillator i, w_i its natural frequency taken from some probability density function $g(w)$, K_{ij} the coupling strength between oscillator i and j (where $K_{ij} \neq K_{ji}$ for unequal coupling), and $\sin(\theta_j - \theta_i)$ is a sinusoidal pairwise interaction function between two oscillators, over which we sum for all coupled neighbours.

2.3 Circadian Phase Response Dynamics

An important question we may ask of any oscillatory system, theoretical or otherwise, concerns how the system responds to perturbations in its rhythm.

2.3.1 Phase Response Curves

Recall the limit cycle oscillator view of circadian clocks, where small perturbations are tolerated by the system, which always relaxes back to some steady state behaviour. This characteristic has in fact already been illustrated in our example of the Goodwin oscillator, which began at some initial quantities of X, Y, and Z (presumably not in equilibrium proportions), and then proceeded toward constant amplitude and period (Fig. 2.2 C).

Perturbations to the circadian limit cycle are ultimately expressed purely through changes in the phase of the oscillation. The Goodwin oscillator model can again be used to illustrate this point (see Fig. 2.5). The response of the system depends on the phase it is in at the point of perturbation, leading to either an advance, a delay, or no change at all. By routinely perturbing circadian clocks at different points along their cycle, we can quantify the resulting change in phase, and construct a 'phase response curve' (PRC) to gain a more complete picture of the response dynamics of the system (Fig. 2.5). For example, it is the comparative ease with which human circadian clocks phase-delay that explains why it is quicker to overcome jet-lag following westward travel (Eastman and Burgess, 2009).

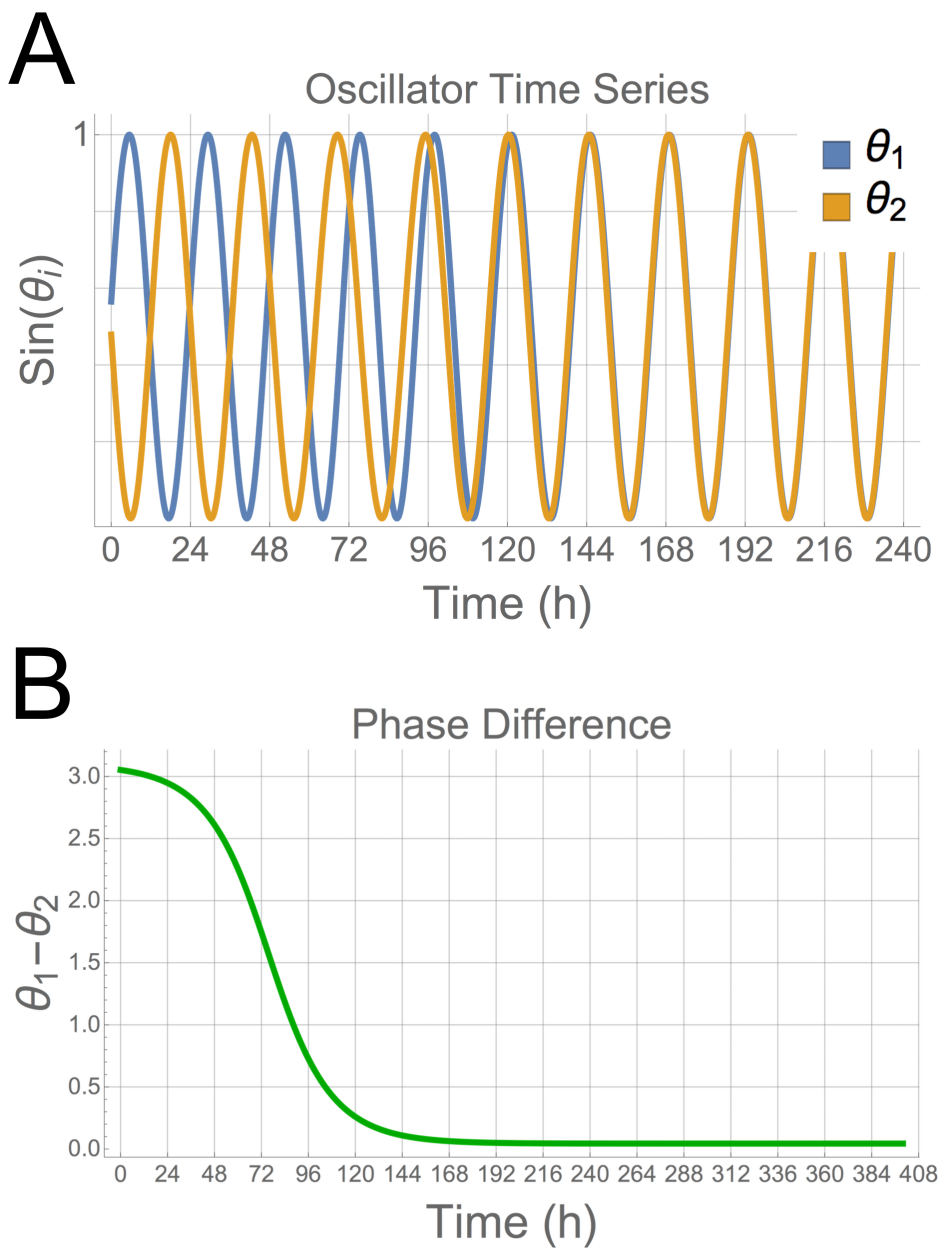

Figure 2.4: Kuramoto phase oscillator simulations. A simple two-oscillator network model illustrating the principles of the WCO framework. Free running period of oscillators are circadian ($w_1 = 24.1hr$ and $w_2 = 23.9hr$) and inter-neuronal coupling is asymmetric ($K_{12} = 0.02$ and $K_{21} = 0.03$). The oscillators begin in antiphase to one another, and gradually align their rhythms as time goes on towards some intermediary phase and free running period (weighted toward oscillator 1 due to the asymmetric coupling). All simulations conducted in *Mathematica*.

2.3.2 Current Challenges

For decades, researchers have sought to map the daily patterns of light-responsiveness in circadian clocks, constructing PRCs as early as 1958 (Hastings and Sweeney, 1958), and leading to a wealth of literature on the topic (for review see Johnson 1999). For the mathematical models, construction of the PRC is a fairly trivial extension (see Fig. 2.5). Indeed, herein lies a strength of *in silico* experimentation. However, the case for real-world biological systems remains more difficult. Let us consider the case of constructing a PRC from circadian locomotor activity rhythms.

The first step is to fully entrain the activity of a population of individuals, using, for example, a 12:12 LD stimulus as we've seen before. The animals are then exposed to constant conditions, say DD, during which a single perturbing light pulse is administered at a given position along the activity cycle. The animals are once more allowed to free-run and activity rhythms stabilise. Finally, the phase of the locomotor rhythms is quantified and compared either to that predicted by the free-running period, or to some un-pulsed control group. The whole process is then repeated for multiple stimulus-delivery times to eventually construct a PRC of the kind shown in Fig. 2.6.

The reader will appreciate from the above protocol that deriving PRCs experimentally is a tremendous undertaking. Indeed, a curve with only 6 data points (denoting the phase shift caused by stimuli delivered at 6 different points in the circadian cycle) can easily take months to construct, and require hundreds of animals. This poses a significant barrier to experimental investigation, and can encourage low n-numbers and resolution in those experiments that are conducted.

In addition to the demand on resources, there is another issue to consider when investigating phase response dynamics from behavioural data – it relies on a suitable method for measuring the phase of the rhythms. This task is non-trivial, as outlined in Section 2.1.

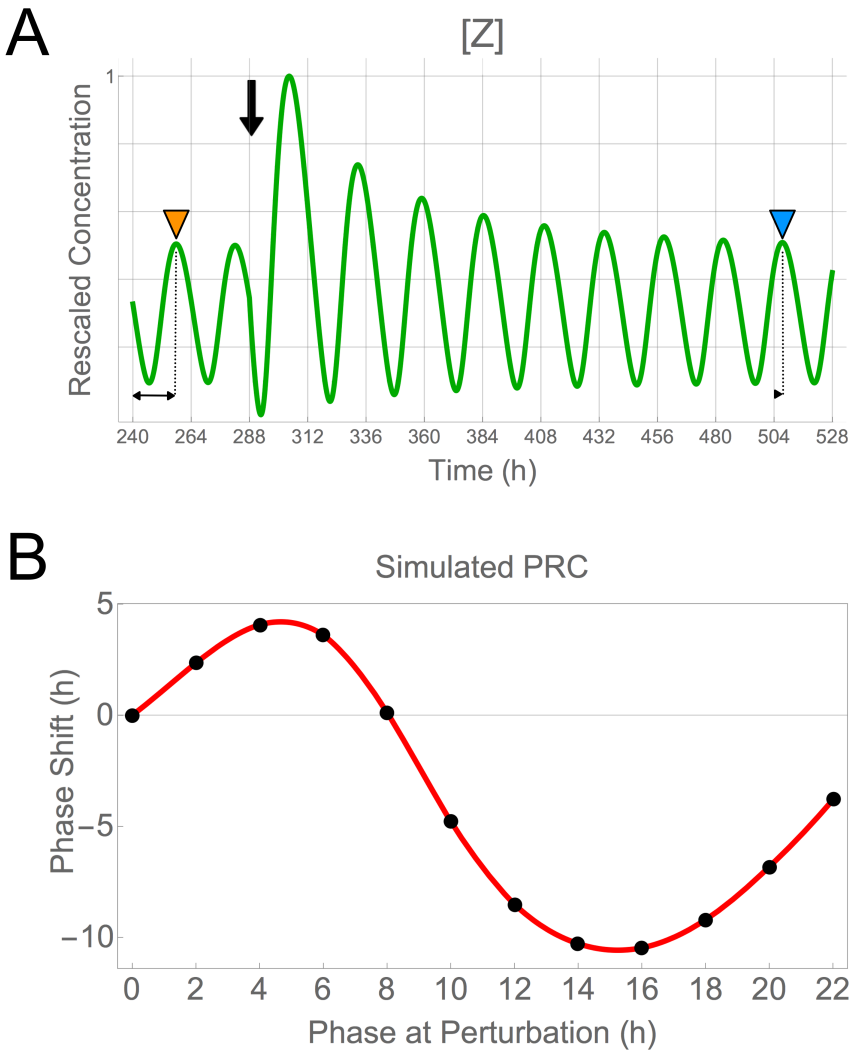

Figure 2.5: Phase response curve derived from Goodwin oscillator model. (A) Concentration of molecule Z from the model defined in Fig. 2.2 plotted in time. The system is perturbed away from steady state at some specified point along the cycle (black arrow) and then relaxes back to equilibrium. Peak phase is measured before (orange triangle marker) and after (blue triangle marker) the perturbation and the phase difference calculated. (B) Phase responses obtained via the process described in (A) plotted against phase along the cycle when the perturbation was administered (black data points). An interpolating function shows the shape of the PRC (red line). We see that the perturbation acts to delay or advance phase of oscillations when applied early or late in the cycle respectively. All simulations conducted in *Mathematica*.

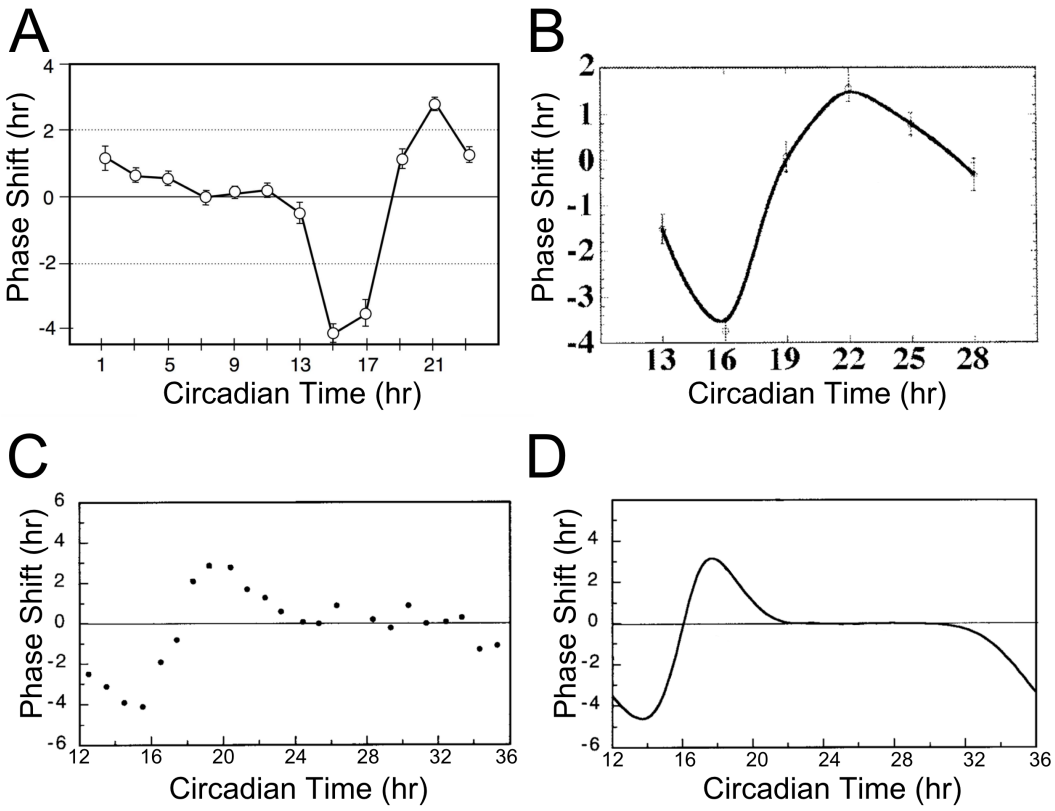

Figure 2.6: Experimental locomotor phase responses to light. (A) PRC showing the average shifts of $12 \times 16 = 192$ wild type flies (w^{1118}) resulting from a 1 hr light pulse delivered at the circadian times indicated (units given in hours, where CT 0 denotes the onset of activity in the morning) Adapted from Vinayak et al. (2013). (B) PRC showing the average shifts of $13 \times 6 = 80$ wild type flies resulting from a 10 min light pulse. Adapted from Stanewsky et al. (1998). (C) PRC showing phase shift of wild type flies in response to a 1 min light pulse. Adapted from Leloup and Goldbeter (1998). (D) Simulated PRC from the mathematical model developed in Leloup and Goldbeter (1998). Note in A, C, and D the presence of a 'dead zone' where external perturbation has little-to-no effect on circadian phase.

Chapter 3

Project Aims & Objectives

The *Drosophila* circadian clock has proven to be an intricate and highly influential neurobiological system, capable of processing multiple sensory signals to drive complex temporally-coordinated outputs. Formative research in the last century has successfully uncovered many secrets of biological timekeeping. However, the field is limited; both by its traditional unimodal approach to investigation, and by a relative lack of mathematical innovation. The overarching goals of my thesis are therefore five-fold:

1. To conduct a systematic study of circadian locomotor behaviour under differentially phased light and temperature cycles. This will provide a more complete assessment of clock function during sensory conflict, and will serve to place previous results in a wider context. I will also seek to explore the pervasive notion of photic dominance, challenging this dogma in a range of laboratory conditions beyond those typically used by researchers (Chapter 5).

2. To link behavioural responses of the clock network, to their underlying molecular oscillations, during conflicting light and temperature input. This will provide a more complete view of sensory processing pathways within the circadian system, as well as the relationship between internal and external representations of the time of day (Chapter 5 and Chapter 6).

3. To conduct the first study of multisensory processing in the peripheral clock network of *Drosophila melanogaster*. This will provide a more hollistic view

of the circadian system, and draw important comparisons between central and peripheral clock function. Greater insight might also be gained through this study on the multisensory properties of mammalian clocks, which share principles in common with peripheral clocks in fruit flies (Chapter 6).

4. To gain further insight on peripheral clock network architecture, investigating the prospect of circadian oscillators in mechanosensory areas of the fruit fly. This would contribute to a fuller picture of multisensory processing by the clock, and introduce the potential for ambiguities between input and output of the system (Chapter 7).

5. To develop and prototype a statistical model of the circadian clock, using techniques drawn from the field of machine learning. This addition to existing mathematical models in the field would be grounded in a more explicit probabilistic representation of the system. Importantly, my modelling approach will seek to provide novel methods for the analysis of circadian locomotor data; with special attention given to statistical inference of clock phase, and the ability to extract richer phase response dynamics from small datasets than is currently possible, together reducing the demand on experimental approaches (Chapter 8).

This thesis will seek to employ a truly inter-disciplinary approach to the neurobiological questions at hand. I aim at all times to maintain a reciprocal flow of ideas and information between the concurrent experimental and theoretical streams of my research.

Chapter 4

Experimental Methodological Overview

This chapter is concerned with the discussion of key methodologies used in my experimental research. As the author of this thesis, and a consumer of scientific literature, my personal preference is to address experimental protocols within close proximity of the results to which they pertain. I therefore adopt this same convention here − materials and methods can be found in their relevant chapters.

The purpose of this chapter, then, is to separately discuss selected items that carry particular significance to my PhD. This includes methods with universal application, spanning multiple chapters of this thesis, as well as analysis tools that required sufficient development as to constitute an independent line of investigation altogether, worthy of reporting.

4.1 Locomotor Assay

Daily rhythms of locomotor activity in *Drosophila melanogaster* permit the investigation of circadian clocks from a behavioural perspective. This powerful experimental tool has contributed to many seminal works in chronobiology, and is accordingly used widely throughout this thesis. Owing to the importance of this protocol, and to remove redundancy for the reader, I here provide details on the locomotor assay used.

4.1.1 Activity Monitoring

The most widely-used locomotor assay is the *Drosophila* Activity Monitoring (DAM) system produced by *Trikinetics*. This automated system allows for continuous and high-throughput data collection over several days. Flies were anaesthetised using carbon dioxide and individually housed in glass tubes containing 5% sucrose and 2% agar medium, which occupied approximately one third of the tube. These tubes were then loaded into activity monitors (*Trikinetics*, Waltham, USA) where infrared (IR) beams are directed through the tubes to a detector, such that 'activity events' are measured each time a fly moves across the beam path (see Fig. 4.1). These events are then summed over the course of the experiment to generate locomotor time series of the kind shown in Fig. 1.5.

At the start my PhD, *Drosophila* circadian behaviour was typically measured using the DAM2 activity monitor (*Trikinetics*; see Fig. 4.2). This 32-channel monitor comprises 2 IR beams directed through the glass tubes at a single point on the mid-line. The construction of these DAM2 monitors seemed to suggest that there would be large parts of an experiment where the fly was moving in the glass tube, but not at the mid-line point, causing the data-collection system to miss this behaviour.

At the time, an alternative multi-beam MB5 monitor that was newly available (*Trikinetics*; see Fig. 4.2), which boasted 17 independent IR beams directed along the length of each tube, raising the question of whether these monitors might be more appropriate for an in-depth numerical analysis of circadian locomotor behaviour (of the form required by this thesis). To address this, I performed a pilot experiment, assessing locomotor rhythms of wild type Canton S flies during aligned LD:TC and 12 hr misaligned LD:TC. This early experiment also served to validate original results on sensory conflict in the fly clock, cf. (Yoshii et al., 2010).

It is clear from the data in Fig. 4.3 that the locomotor waveforms produced by the DAM2 and MB5 monitors are different. The DAM2 monitor does indeed underestimate fly activity, with maximum activity measured by the MB5 to be over ten times greater. We can also see greater noise in the histograms produced by

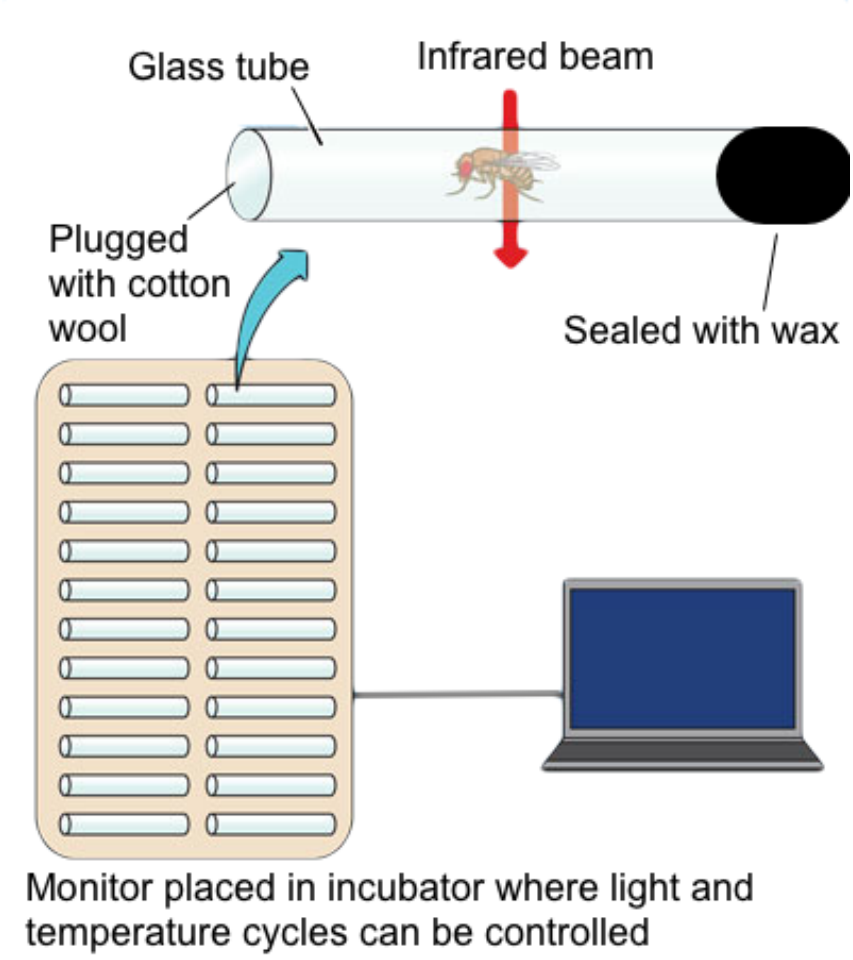

Figure 4.1: Measuring locomotor activity in *Drosophila melanogaster*. Experimental set-up showing individual flies placed inside glass activity tubes, which are then loaded into *Trikinetics* monitors where infrared beams are directed through the tubes to measure activity levels (see text). This experimental set-up is typically placed inside incubators to establish control over environmental conditions.

DAM2 data, reflecting the lower resolution measurements for this device (Fig. 4.3). Furthermore, the effect of a 12 hr conflict between LD and TC appears more pronounced in the MB5 data, which shows a broader, drawn out peak with earlier onset of evening anticipation behaviour, as has been reported previously (Fig. 4.3; Yoshii et al. 2010). Curiously, the DAM2 data appears to show the reverse, with evening anticipation onset starting later in the conflicting conditions. This could perhaps be a result of noise in the data.

For the reasons outlined above, more MB5 monitors were acquired, and, unless otherwise stated in the text, all locomotor assays presented in this thesis have been

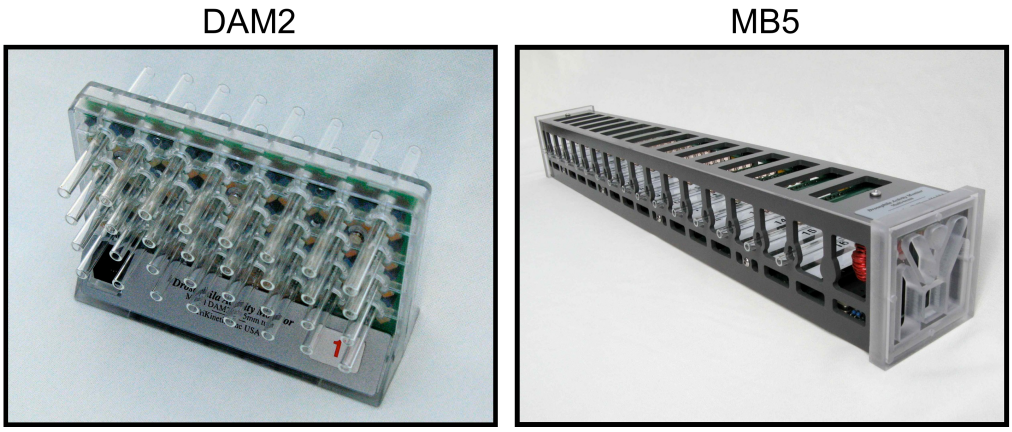

Figure 4.2: *Trikinetics Drosophila* activity monitors. Photographs of the 32-channel DAM2 monitor with 2 IR beams (left), and the 16-channel MB5 monitor with 17 IR beams. Images taken from the *Trikinetics* website (www.trikinetics.com).

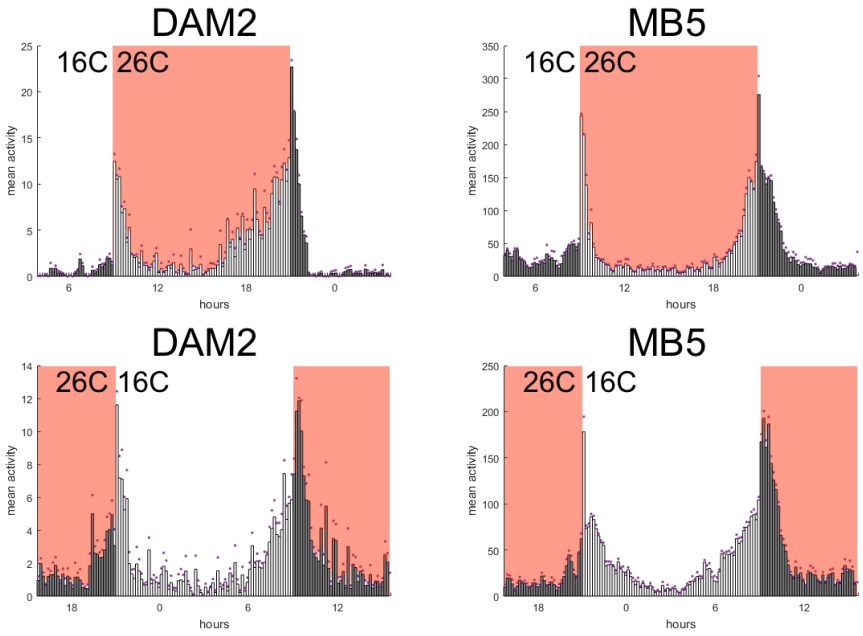

Figure 4.3: Comparison of data collected from DAM2 and MB5 monitors. Pilot locomotor assay using either DAM2 monitors (left) (n = 16) or MB5 monitors (right) (n = 16). Flies were entrained to aligned LD:TC and then subsequently to misaligned LD:TC as was done in Yoshii et al. (2010). Histograms show average activity across two consecutive days in each entrainment condition, with 10 min bin size. Environmental conditions marked clearly on the plots - white/dark bars show light/dark respectively; red/white background shows warm/cold respectively.

conducted using this superior experimental set-up.

4.1.2 Data Analysis

I have used the well-established *MATLAB* flytoolbox library for the analysis of circadian locomotor data in *Drosophila* (Levine et al., 2002a). Activity of individual flies and average activity of the population are typically plotted as 'double actograms' showing a 48 hr time window for each row of the actogram to facilitate uninterrupted visualisation of any 24 hr section of the data. Where data are not plotted in *MATLAB*, care has been taken to highlight this in the text.

4.1.2.1 Data conditioning

In order to extract valuable metrics from circadian time series data, the signal often requires a degree of pre-processing. For the analysis of locomotor activity in the fruit fly, pre-processing includes the use of a digital filter to simultaneously detrend long-term changes and normalise the signal (see Levine et al. 2002a for details).

Briefly, a low-pass Butterworth filter is applied once in the forward direction, and once backward, to mitigate any effect of phase shift in the data. Filter size is set to 72 hr (3×24) to extract the long-term dynamics in the data. Each data point in the raw trace is then divided by its corresponding point on the trend line to produce a normalised trace with mean 1 and long-term trend removed. Analysis is then performed on this conditioned signal.

4.1.2.2 Autocorrelation

Autocorrelation is a powerful tool to extract periodicity and rhythm strength from circadian time series data. It measures the degree of similarity between a given time series and a lagged version of itself over successive time intervals. Put another way, autocorrelation describes the relationship between a variable's current value and its past values. Naturally, if the data displays oscillations, the correlation between a given point and lagged points will show high correlation at times when the lag-size corresponds to the natural frequency, or period, of the oscillations.

Using the *MATLAB* flytoolbox library, autocorrelation is used to generate a correlogram (Fig. 4.4), for which the location of the peaks reveal the period of cir-

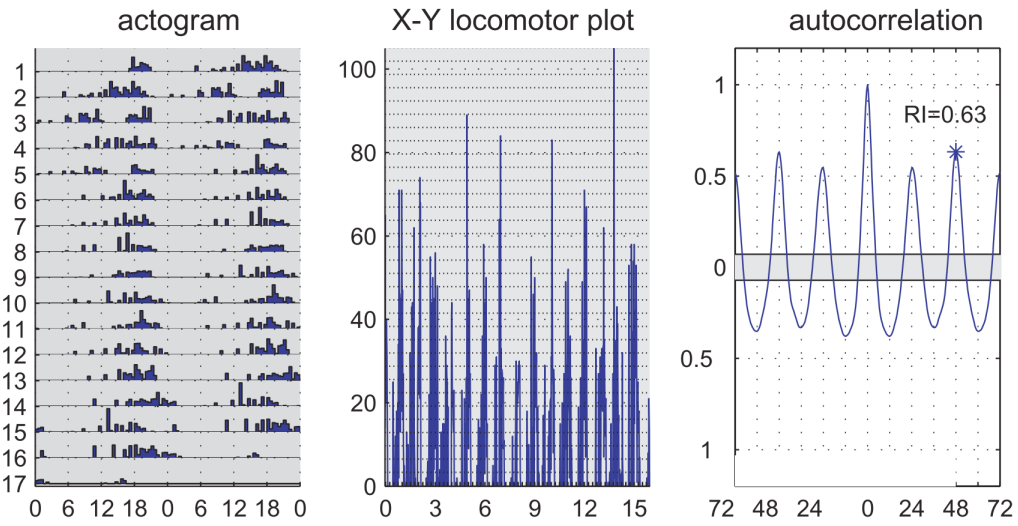

Figure 4.4: Autocorrelation analysis of fly locomotor behaviour. Free running locomotor rhythms (DD) are plotted either as an actogram (left) or against time (middle). Autocorrelation is run on the data to produce a correlogram (right), where the period is shown to be 24.5 $\pm$ 0.1, and RI shown by the blue asterisk. Adapted from Levine et al. (2002a).

cadian oscillations present in the data (Levine et al., 2002a). To specify the *strength* of the rhythms, a rhythmicity index (RI) is calculated as the height of the third peak in the correlogram (Levine et al., 2002a), and a corresponding rhythmicity statistic (RS) is calculated as RI divided by the 95% confidence line in the correlogram, thus describing statistically significant rhythms (Levine et al., 2002b).

4.1.2.3 Phase

As described in Chapter 2 (page 40), phase analysis of circadian time series data is non-trivial. With slightly less restrictive assumptions and an ability to handle non-stationary data, I chose to implement a non-parametric approach for the phase analysis of locomotor rhythms, using computer functions adapted from the *MATLAB* flytoolbox library (Levine et al., 2002a).

For individual flies, data averaged over two or more consecutive days (beginning at some reference time; typically ZT 0) is smoothed using a low-pass Butterworth filter of width = 12 hr (see Fig. 4.5 for example of the filtering procedure). The major peaks are located, and peak phase relative to the reference time is calculated in hours (Fig. 4.5). The results of this analysis are typically plotted in polar

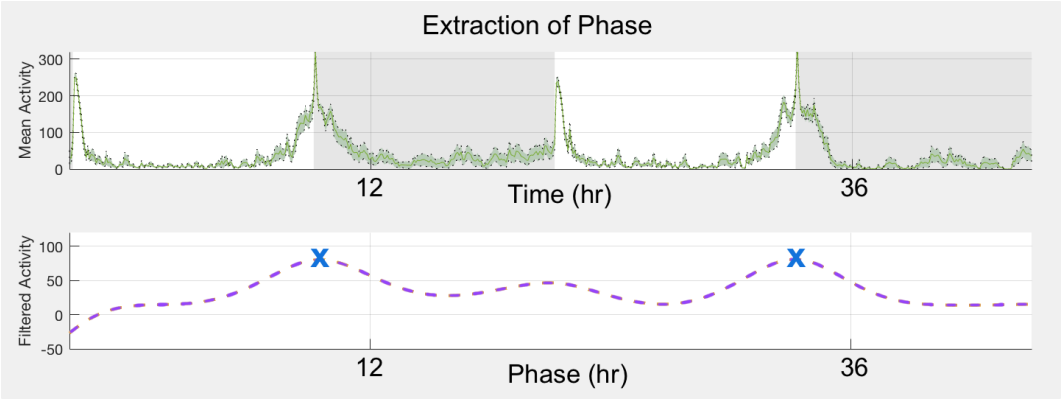

Figure 4.5: Butterworth filter applied to fly locomotor data. A forward-backward Butterworth filter (filter size = 12 hr) is applied to two days of wild type locomotor behaviour during aligned 12 hr: 12 hr LD:TC (Top; data taken from Fig. 4.3, top-right panel; n = 16). Major peaks are located and peak phase evaluated (Bottom; blue crosses show major peaks).

coordinates and, when comparing between two groups, a dispersion test is used to determine whether the two distributions differ significantly in angular deviation from their respective means (see Batschelet 1981, for full review).

4.2 Luciferase Assay

In recent years, real-time bioluminescence recording using firefly *luciferase* as a reporter has contributed greatly to the study of circadian rhythms. The development of transgenic reporter constructs, together with luminescence detection apparatus, has enabled high-frequency monitoring of circadian clock function in living tissues (Brandes et al., 1996; Stanewsky et al., 1997b, 1998, 2002; Veleri et al., 2003).

Strategies for the analysis of bioluminescence data are outlined in Levine et al. (2002b) and Levine et al. (2002a). However, there exists no widely used software implementation of these strategies beyond individual functions in the wider *MATLAB* flytoolbox library. To improve on existing methods for the analysis of *luciferase* reporter data, and integrate this in a form that is accessible to the circadian researcher, I developed a new library to sit on top of the flytoolbox library – hereafter referred to as *Luceo* (from the same Latin root as *luciferase*, and meaning 'to be clear', in-keeping with its intended purpose to simplify data analysis).

The *Luceo* toolbox serves to formalise the strategies described in Levine et al.

(2002a), while offering additional advantages described below.

4.2.1 *Luceo*: A Custom Toolbox for the Analysis of Bioluminescence Rhythms

In Chapter 6 of this thesis, I use the *luciferase* reporter assay to investigate peripheral clock function in the *Drosophila* circadian system. Thus, in an effort to remove redundancy, I here provide details of the analysis protocol used in Chapter 6, providing details relevant to the experiments conducted in Chapter 6, as well as presenting the reader with a complete account of the *Luceo* work flow. All analysis was conducted using the *Luceo* bioluminescence toolbox.

First, raw bioluminescence data was loaded into the *MATLAB* environment and individual traces plotted to permit visual inspection of the rhythms. As flies can perish during the course of an experiment, *Luceo* allows the researcher to select sections of the data prior to death for use in subsequent analysis (Fig. 4.6). This prevents unnecessary data loss from automatically excluding all flies that die (even if death occurs in the last days of an experiment). Note that n-numbers within a group can accordingly vary between different regions of analysis during a single experiment (reported in the text).

A mean trace of ~20-40 tim^{01}*;XLG-luc1-1* control flies lacking a functional clock was subtracted from each individual in the *XLG-luc1-1* and *XLG-luc1-1* cry^{b} test groups to remove components of the bioluminescence signal that were not clock-driven (Fig. 4.7; see discussion on page 92). Using the filtering method described previously for locomotor data (page 60), individual traces were detrended and normalised to have a mean time course equal to 1, and preserve the appearance of percentage changes for oscillations around the long-term trend line (Fig. 4.7).

Individual bioluminescence signals in the *XLG-luc1-1* and *XLG-luc1-1* cry^{b} groups were next quantified to produce an empirical distribution of values, from which the median and 95% confidence intervals were calculated (avoiding the assumptions of normality associated with the mean average). Non-parametric Mann-Whitney U tests were used for statistical comparisons between genotypes and environmental conditions.

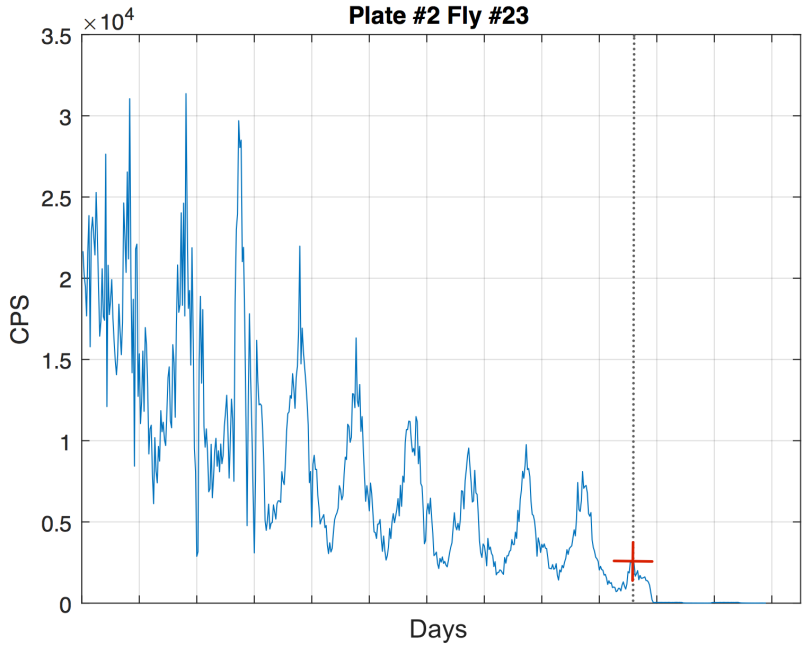

Figure 4.6: Example of data selection in *Luceo*. The researcher can manually include data up to a certain point by dragging the crosshair marker to the desired location. This process can also be automated according to activity threshold rules.

Peak phases were determined by first smoothing individual traces using a low-pass Butterworth filter to remove noise (filter size = 12 hr). The mean peak phase was then measured across multiple days according to the reference time, ZT_T = 0 (Fig. 4.7). The amplitude was measured from the smoothed trace as the mean difference between the peak and trough across multiple days (Fig. 4.7). Finally, autocorrelation analysis of the raw signal was used to calculate the rhythm statistic (RS) as the height of the third peak in the correlogram divided by the confidence interval (described on page 59; Levine et al. 2002b). As autocorrelation is sensitive to the length of the time series provided, exactly two days were used for all analyses of this type. Free running period was estimated as the location of the third peak in the correlogram divided by two (Levine et al., 2002a). A mean average trace, with shaded region showing SEM, was then plotted for each genotype using the ggplot package in *R* (Fig. 4.7).

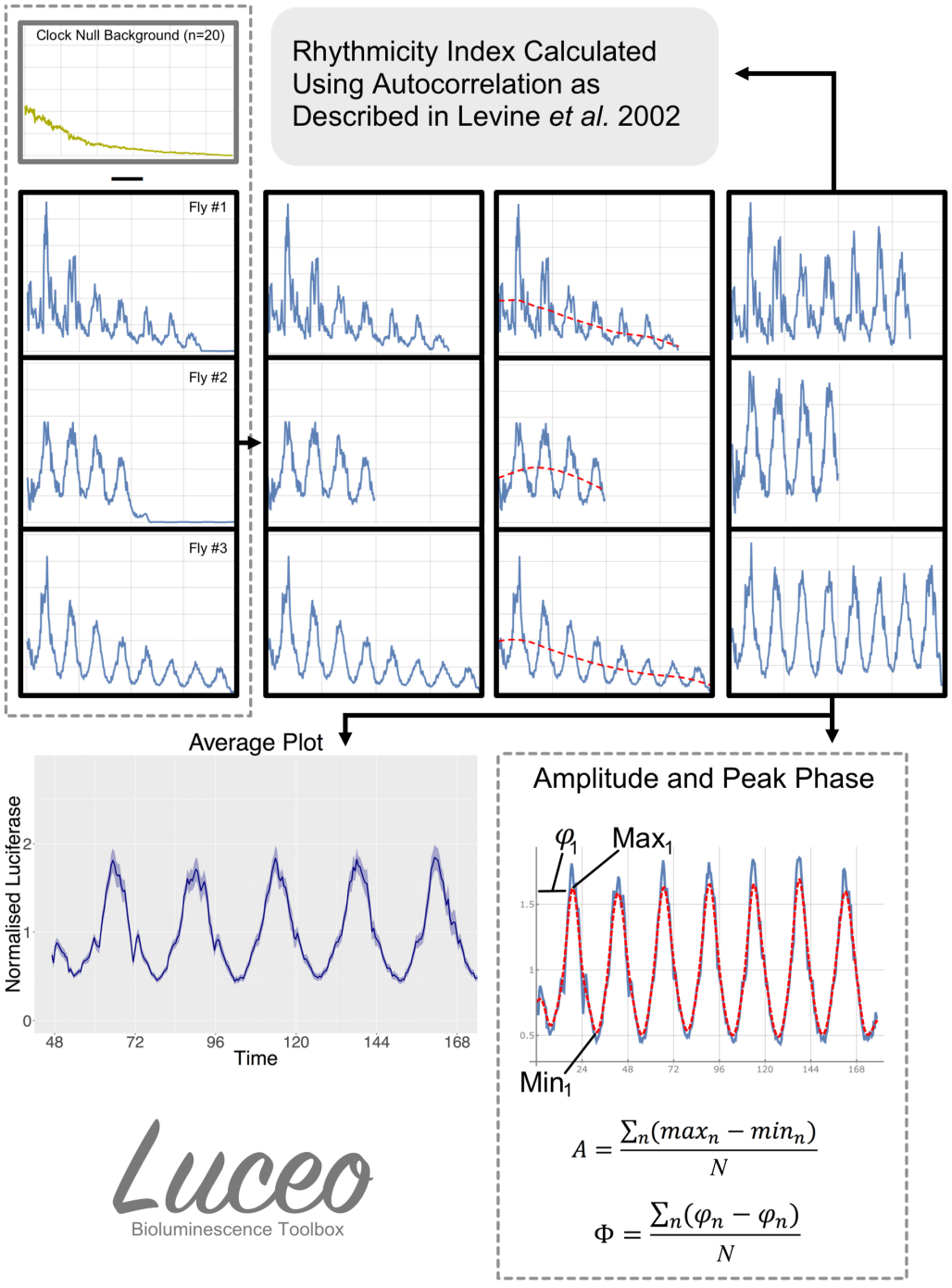
Clock Null Background (n=20)
Rhythmicity Index Calculated Using Autocorrelation as Described in Levine et al. 2002
Fly #1
Fly #2
Fly #3
Average Plot
Normalised Luciferase
Time
Amplitude and Peak Phase
φ_1
Max_1
Min_1
$A = \frac{\sum_n (max_n - min_n)}{N}$
$\Phi = \frac{\sum_n (\varphi_n - \varphi_n)}{N}$
Luceo
Bioluminescence Toolbox

Figure 4.7: Bioluminescence data analysis work flow in *Luceo*. Raw data from individual wells (flies) were extracted. A mean average trace of a clock-null mutant (*tim^{01};XLG-luc1-1*) was obtained, and this was subtracted from each individual in the test genotypes (*XLG-luc1-1* and *XLG-luc1-1 cry^{b}*). Where emissions drop to background levels (50-100 CPS) flies were deemed to have died and data from this point onwards were excluded from subsequent analysis. Each individual signal was then detrended and normalized as described in the text by applying a low-pass filter to extract the long-term trend and dividing each point in the raw trace by this baseline. Peak phase, amplitude, and rhythmicity were then measured from these individual normalized signals to generate an empirical distribution of these metrics over the population, and finally a mean average was plotted in *R* with shaded region showing SEM. All analysis conducted in *MATLAB* with final plots generated in *R*'s ggplot package.

The *Luceo* toolbox offers a dedicated library for the analysis of bioluminescence data, making use of currently leading methods in the field (including the autocorrelation, detrending, and normalisation described in Levine et al. 2002a). The toolbox permits convenient visualisation of the data, non-parametric statistical tests, and exports the processed data as a 'dataframe' structure for preparing images using the excellent ggplot package in *R*. Unlike existing software implementations, *Luceo* allows for the easy selection of data up to the point of animal death, increasing the analytical power for longer experiments, where viability issues can arise. Furthermore, the toolbox provides a method to subtract from the test-group data, traces from a control-group (here circadian mutants were used to remove non-circadian components from bioluminescence traces).

The *Luceo* toolbox is undergoing development from a library of functions, to a graphical interface integrating the above functionality into one easy-to-use environment. It is expected to be made available later this year.

Chapter 5

Sensory Conflict Disrupts Activity of the *Drosophila* Central Clock Network

The results presented in this chapter have been published in *Cell Reports* (see Harper et al. 2016)

5.1 Introduction

In Chapter 1, page 33, we discussed the small number of previous studies that have begun to embrace a more multisensory treatment of the *Drosophila* circadian clock. These studies used misalignments between the daily cycles of light and temperature to construct sensory conflicts in the environment. Function of the circadian system in response to these conflicting sources of information was then assessed through analysis of locomotor behaviour or molecular cycling in the brain.

Interpretation of these previous experiments has generally included some notion of light dominance, which fits nicely with a prevailing view in the field of chronobiology. However, we cannot ignore that temperature also appears to exert circadian effects on the timing of evening anticipatory behaviour during misaligned light and temperature cycles. The situation thus remains unclear.

Let us now review these previous experimental results through a more theoretical lens, using the Bayesian framework outlined in Section 1.4.2.

At its core, the clock network must perform multisensory integration (MSI). In the Bayesian characterization of timekeeping, there is a hidden or latent variable (here, the true time of day) whose values are associated with possibly noisy observations (fluctuating light and temperature signals). In the context of previous results, Bayesian analysis brings to the fore two key considerations: 1) the relative strengths of different signals, and 2) the possibility that the signals might have different, as opposed to the same, underlying causes.

As we saw in Chapter 1, different sources of an observation can be integrated with different weights of influence according to their respective reliabilities. Weak periodic fluctuations in a Zeitgeber will thus provide little reliable evidence about the time of day, and so exert little effect over the estimate. In the study by Miyasako et al. (2007), sensory misalignment was generated through a 6 hr advance of TC relative to LD. The authors of this study found locomotor rhythms largely entrained to the light cue, implicating this as the dominant signal for circadian locomotor entrainment in *Drosophila melanogaster*. However, in this study, temperature oscillated between 20°C and 25°C. Our theoretical reasoning suggests that this comparatively small amplitude TC, for what is regarded as the weaker of the two ZGs in *Drosophila* (Yoshii et al., 2010), may have been insufficient to distinguish subtle signal averaging effects from background noise, especially given the much larger temperature ranges found in nature (Vanin et al., 2012).

Bayesian treatments of MSI also acknowledge the possibility that highly discrepant signals are unlikely to come from the same underlying value of the latent variable (Körding et al., 2007). Depending on the circumstance, inference could then reject one of the signals as being just noise; or it could infer that there is more than one underlying latent variable (Fig. 5.1 right). In these cases, the smaller the disparity between the signals, the readier inference will be to integrate them (Fig. 5.1 left).

In the study by Yoshii et al. (2010), LD and TC were misaligned by 12 hr − an antiphasic relationship that represents the largest possible disparity between two 24-hr environmental oscillators. During this extreme sensory conflict, activity rhythms

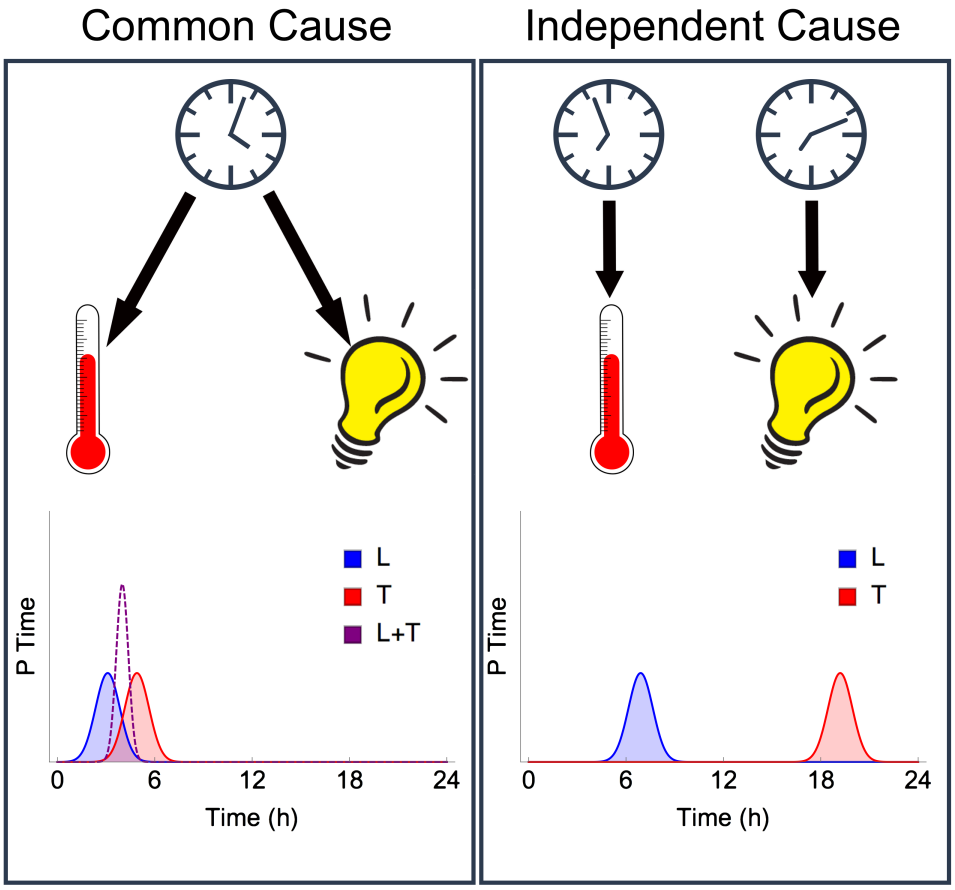

Figure 5.1: Causal inference in the circadian clock. A diagram depicting the challenges of inferring causal structure in the world. From the observable variables, light and temperature, the circadian clock must infer the time of day. When the times suggested by the observable variables are similar (left), Bayesian integration would combine these cues to yield a more confident estimate. In the example above, equal weighting is given to light and temperature. When the times suggested by the observable variables are highly discrepant (right), the clock network might infer a causal structure where there are two different times of day.

of wild type flies entrained preferentially to the light stimulus, leading the authors to the conclusion that the light cue is dominant. However, under such extreme conflict, it is possible that two different times of day were inferred by the clock, one each associated with light and temperature (as in Fig. 5.1 right). This would be consistent with the existence of light- and temperature-sensitive neuronal subgroups within the clock network, and may govern a more 'all-or-nothing' entrainment of locomotor rhythms to only one of the cues (in this case, light).

Theoretical review of previous studies highlights aspects of their experimental set-up that may have fallen short of uncovering potentially richer forms of MSI in the *Drosophila* circadian clock. It also seems reasonable to suggest that the analysis of one single signal disparity between cues is insufficient to fully probe the possible coupling at play.

To better understand the effect of environmental phase relationships on circadian clock function, I assessed circadian locomotor behaviour during misaligned LD and TC using finer gradations of sensory conflict (contrasting Yoshii et al. 2010 and Currie et al. 2009) and large diurnal fluctuations in both cues (contrasting Miyasako et al. 2007). Furthermore, like Yoshii et al. (2010), I compared wild type flies to *cry*-null mutants, removing the key contribution made by the circadian photoreceptor Cryptochrome (CRY) to light entrainment of the clock (Stanewsky et al., 1998). I hypothesized that any effect of MSI would be markedly diminished in *cry* mutants, owing to a reduced weight of the light-dependent input pathway and relative enhancement of the temperature cue (Gentile et al., 2013).

5.2 Materials & Methods

5.2.1 Fly Strains

Canton S flies were used as wild type flies. *Cryptochrome* mutants were $w;cry^{02}$ and $w;cry^{01}/cry^{b}$ (Dolezelova et al., 2007; Stanewsky et al., 1998). For rescue experiments, *Clk856-gal4* (Gummadova et al., 2009), and *tim-gal4:27* (Kaneko and Hall, 2000), were crossed into a homozygous mutant cry^{b} background (Stanewsky et al., 1998) using appropriate balancer chromosomes and dominant markers. These

gal4 driver lines were then crossed to homozygous cry^{01} flies carrying *UAS-cry24.5* on chromosome 2 (Emery et al., 1998). First generation *Clk856-gal4/UAS-cry24.5* or *tim-gal4:27/UAS-cry24.5*, cry^{b}/cry^{01} males were then analysed behaviourally as described below.

Flies were reared under 12:12 LD cycles on *Drosophila* medium (0.8% agar, 2.2% sugar-beet syrup, 8.0% malt extract, 1.8% yeast, 1.0% soy flour, 8.0% corn flour, and 0.3% hydroxybenzoic acid) at 25°C and 60% humidity. Only male flies at an age of 3 to 6 days were used in experiments.

5.2.2 Activity Monitoring

Locomotor activity rhythms were recorded automatically using the *Drosophila* Activity Monitoring (DAM) system (*Trikinetics*, Waltham, MA) as previously described in Chapters 1 and 4. Analysis of locomotor data was conducted as described in Chapter 4.

5.2.3 Quantification of Entrained Behaviour

Analysis of locomotor behaviour under entrained conditions is inherently challenging as observed activity must be a result of both circadian drive and direct sensory effects (e.g. startle behaviour and masking; see discussion on page 30). It is common within the field to assess the anticipatory behaviour prior to ZG offset – the so-called 'evening activity'. However, this is usually performed by eye.

I sought to implement a more quantitative approach to quantifying the extent of evening anticipation to a given cue. It seems reasonable to suggest that a simple measure of this evening behaviour is a *linear* increase of activity prior to ZG offset. With this in mind, I used maximum likelihood estimation to fit a linear model ($y = a + bx$) to the activity bout immediately preceding both light and temperature evening (Fig. 5.2). Here, the gradient of the fitted linear model (parameter b) reflects the degree of evening anticipation, with more positive values denoting strong evening anticipation, and less positive/more negative values indicating weak or no evening anticipation. Data points falling on or 1 hr after the environmental transitions were excluded from the analysis due to the confounding effects of behavioural

masking.

In cases where ZG offset for light and temperature were in close proximity, care was taken not to include evening activity for one stimulus in the analysis of the second. Thus, during very small or very large conflicts, fewer data points were available for fitting the later Zeitgber, which translated into larger confidence intervals for these time points.

5.2.4 Immunostaining and Quantification

Flies collected at four time points during the alignment conditions (corresponding to ZT 3, ZT 9, ZT 15 and ZT 21 of the aligned condition) were fixed in 4% paraformaldehyde in 0.1M phosphate buffer (PB) with 0.1% Triton X-100 (PBS-T) for 2.5 hr at room temperature.

Flies were then rinsed three times in PB, and the brains subsequently dissected in PB. Brains were then blocked in 5% normal goat serum (NGS) in 0.5% PBS-T at 4°C for 36 hr before incubation in primary antibodies for 48 hr at 4°C. Double staining was conducted with primary antibodies: rabbit anti-PER (1:1500) (Stanewsky et al., 1997a), and mouse anti-PDF (1:500) (DSHB).

Secondary antibodies were applied after washing six times in 0.5% PBS-T. These were alexaFluor 488 and alexaFluor 647 (purchased from *Thermo Fisher Scientific*), and were both diluted 1:300 in 0.5% PBS-T. After incubation with secondary antibodies, the brains were washed six times in 0.5% PBS-T and mounted in Vectashield (*Vector Labs*) mounting medium. The fluorescence signals of the wholemount brains were visualized using a Leica SP8 laser scanning confocal microscope.

Quantification of PER signals was conducted without discrimination of subcellular localization using the program *ImageJ*, as described previously (Rieger et al., 2006). Briefly, for each neuron, mean pixel intensity of PER staining was measured, apart from the DN3 neurons, which were scored as a group due to their small size and large number. A final mean staining intensity was calculated from all neurons within a group per hemisphere. Three background staining levels were then measured in the surrounding field of each neuronal group and the mean value of

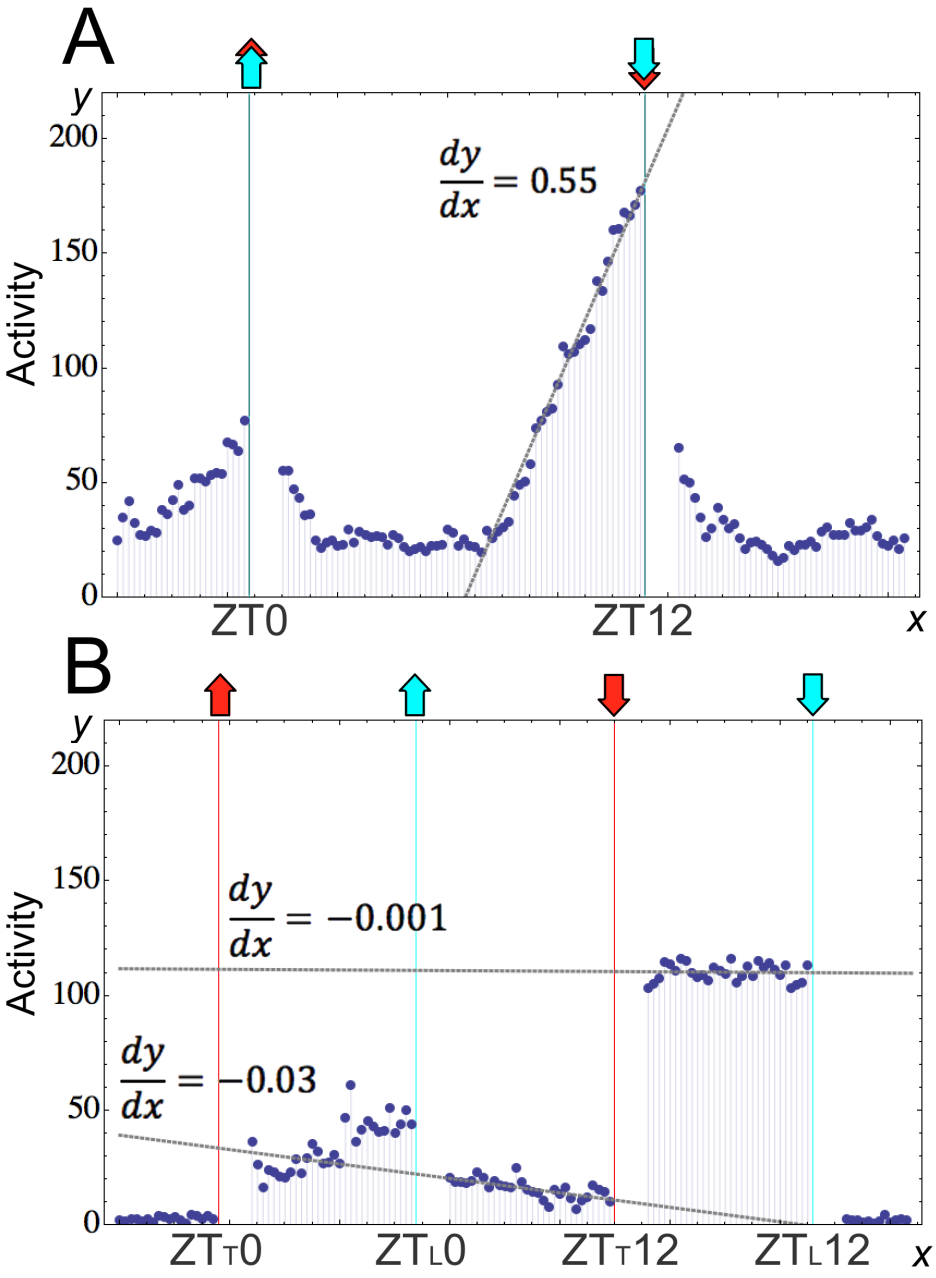

Figure 5.2: Novel method for the quantification of evening anticipation behaviour. Linear regression is applied to the section of activity prior to ZG offset using MLE. Data points at, and 1 hr after, the ZG transitions are excluded from the analysis. (A) When light and temperature are in-phase ($ZT_T = ZT_L$), the linear increase of activity towards ZG offset is comparatively steep (b is large), showing typical locomotor entrainment. (B) When light and temperature are out-of-phase by 6 hr ($ZT_T \neq ZT_L$), we fit two linear regressions: one for the activity preceding temperature offset (red line), and one for the activity preceding light offset (cyan line).

these three measurements was subtracted from the mean pixel intensities measured for the cells.

A final staining index was calculated for each cell group:

$$SI = (\mu_{cells} - \mu_{\varepsilon}) \times \frac{N_{cells}}{N_{max}} \tag{5.1}$$

where SI is the staining index, μ_{cells} is mean pixel intensity for a clock cell group, μ_{ε} is the mean background intensity, N_{cells} is the number of stained cells counted, and N_{max} is the maximum number of cells for the given group. N_{max} for the different neuronal groups was as follows: s-LN_v (4), l-LN_v (5), LN_d (7), DN1 (17), DN2 (2). Again, owing to the large number of DN3 neurons, SI for this subgroup was calculated as group mean pixel intensity minus background.

5.3 Results

5.3.1 Sensory Conflict Disrupts Normal Daily Locomotor Activity

While recent studies have aimed to generate more naturalistic environmental transitions (e.g. Vanin et al. 2012), my study of the mechanistic bases of ZG integration requires the establishment of deliberately *unnatural* experimental conditions. Note that I refer to cue misalignment as the absolute distance, in hours (delta time, or Δt), between onset/offset of two cyclic 12 hr:12 hr signals. For example, $\Delta t_{L,T}$ = 3 hr denotes that light onset/offset occurs 3 hr after temperature.

Wild type and *cry*-null mutant flies were subjected to an environmental regime comprising aligned LD:TC (Part I, $\Delta t_{L,T}$ = 0 hr), followed by a 6 hr delay of LD with respect to TC (Part III, $\Delta t_{L,T}$ = 6 hr), interspersed or followed by free running conditions to assess stability of the endogenous rhythms (Part II and Part IV, outlined in Fig. 5.3 A). As is standard practice for observing endogenous activity rhythms, free running conditions comprised constant darkness and constant warmth (26°C *Drosophila*'s preferred ambient temperature (Sayeed and Benzer, 1996)) to

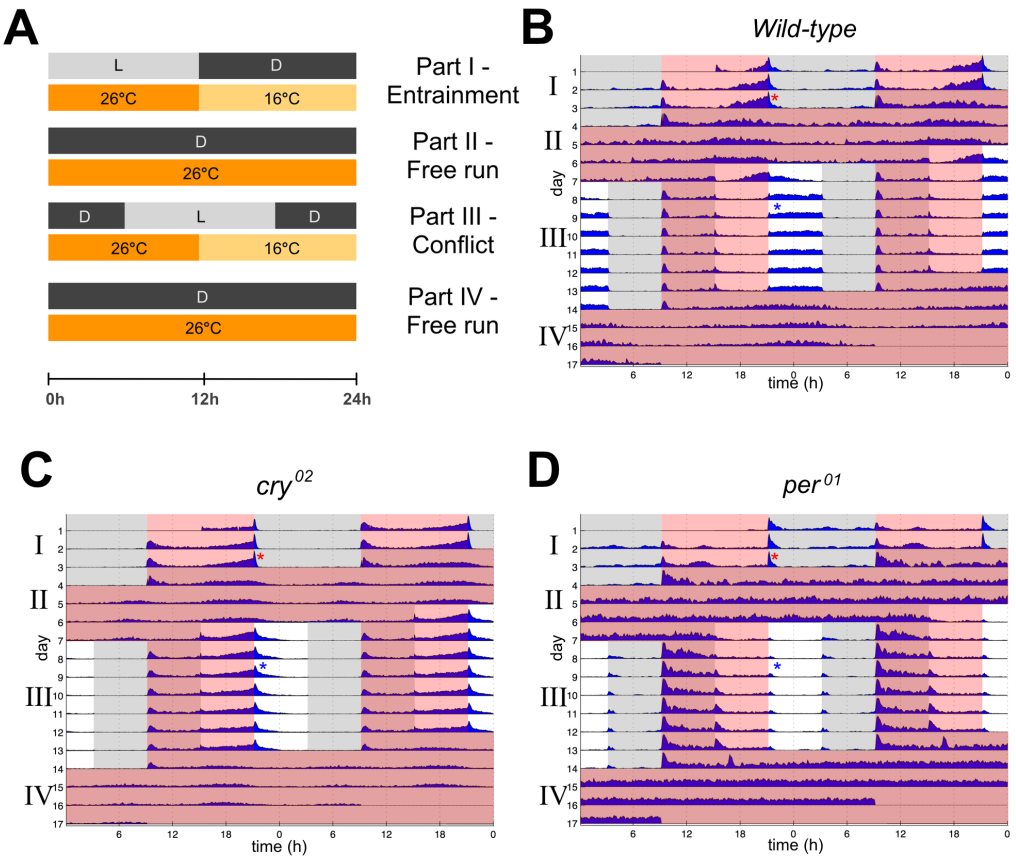

Figure 5.3: Locomotor behaviour during sensory conflict. (A) Experimental regime in which environmental conditions followed 3 days of 12 hr:12 hr LD and TC (16:26°C) in-phase (I), 3 days of free run in DD at 26°C (II), 7 days of out-of-phase 12 hr:12 hr LD and TC (16:26°C) via 6 hr delay of LD (III), followed by 3 days of free run in DD at 26°C (IV). (B-D) Average actograms of wild type (B) (n=46), cry^{02} (C) (n=44) and per^{01} (D) (n=32). Red asterisk denotes representative evening behaviour in Part I; blue asterisk denotes representative pseudo-evening behaviour in Part III. Clock-less per^{01} flies show only brief startle responses to the sudden environmental changes and otherwise display arrhythmic behaviour (C). See Fig. 5.4 for individual fly data, and Fig. 5.5 for controls.

mitigate any negative masking effect of cold temperatures on overall activity levels.

In Part I, locomotor behaviour in wild type and cry^{02} flies both displayed a characteristic bimodal profile, showing an evening peak of activity that coincided with the end of photo/thermo-phase (Fig. 5.3 B,C). These entrained rhythms persisted in free-running conditions (Part II). In Part III, a 6 hr misalignment between LD and TC was introduced via a 6 hr delay of LD relative to Part I (leaving TC un-

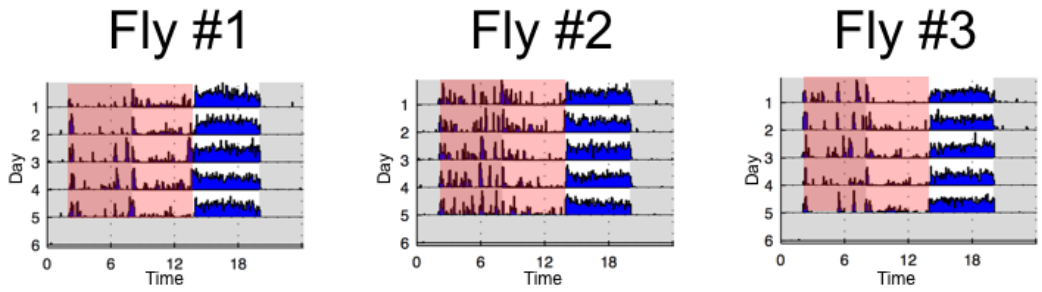

Figure 5.4: P behaviour observed for individual flies. Three representative activity traces taken from individual Canton S flies during 6 hr misalignment of LD and TC (16:26°C). Red filter denotes warm phase; white background denotes light phase.

changed). During this sensory conflict, circadian locomotor behaviour in wild type flies was drastically altered, exhibiting a plateau of sustained activity between temperature offset and light offset, bordered by periods of inactivity (Fig. 5.3 B). The activity pattern continued for the duration of the conflict and was also seen at the level of individual flies, and across multiple repeats (Fig. 5.4). It was also observed for multiple experimental repeats (Fig.5.5 E).

A key facet of the novel activity pattern is the absence of any evening anticipation to either the light or the temperature cue. For ease, I refer to this abnormal locomotor behaviour as Plateau (P) behaviour. Importantly, P behaviour depends on a functional clock as it cannot be observed in per^{01} mutants (Fig. 5.3 D). That P behaviour is not merely induced by masking is also apparent from comparing the free running behaviour in Part IV with that in Part II (Fig. 5.3 B; Fig. 5.5 E).

The P behaviour observed in wild type flies was not present in cry^{02} mutants during conflict conditions, which instead displayed the typical ramping increase of activity, peaking at temperature offset (Fig. 5.3 C). This suggests these flies predominantly entrained to TC. However, the reader will note the behavioural profile is slightly altered from that in Part I, for instance including an extended period of activity after temperature offset. Indeed, the conflicting regime (and therefore the periodic presence of light) appears to have had some effect, albeit greatly reduced, on the behaviour of cry^{02} mutants. This observation is consistent with the existence of *cry*-independent light entrainment pathways (Yoshii et al., 2016).

To test whether the absence of P behaviour in cry^{02} mutants was indeed due to the absence of CRY, I rescued *cry* expression in all clock cells or all clock neurons (*tim-gal4/UAS-cry;*cry^{b}/cry^{01} and *Clk856-gal4/UAS-cry;*cry^{b}/cry^{01}, respectively). Rescue flies displayed activity rhythms that more closely resembled the wild type than the cry^{02} pattern – inactive prior to temperature offset, with a constitutive bout of activity between temperature and light offset (Fig. 5.3 B; Fig. 5.5 A,B,E). These data suggest that it is indeed the *integration* of two potent, yet conflicting input signals to the clock – one photic and the other non-photic – that underlies the abnormal behavioural output observed during sensory conflict.

5.3.2 Sensory Conflict Disrupts Endogenous Oscillations in the Central Clock Network

On page 27 of Chapter 1, I discussed how cytological staining for clock gene products has revealed the location of the central circadian network in *Drosophila*, which can be further classified into seven distinct cell groups: s-LN_vs, l-LN_vs, LN_ds, DN1s, DN2s, DN3s, and LPNs. While there are likely to be additional subdivisions within the network (Peschel and Helfrich-Förster, 2011), my study of multisensory processing in the fly brain will adopt the current prevailing, and well-supported, network architecture. The reader will recall at this stage that neurons in the central clock network appear to exhibit preferences for light and temperature (Yoshii et al., 2010).

To examine the molecular and neuronal substrates of the pronounced P behaviour, I carried out antibody staining for the clock protein PERIOD (PER) in the *Drosophila* brain during 6 hr misaligned LD:TC (the same degree of conflict that generates P behaviour). PER immunostaining of clock neurons was performed at four time points evenly spaced across 24 hours (ZT 3, ZT 9, ZT 15, and ZT 21) for both Part I and Part III of the experimental regime described in Fig. 5.3 A. In flies that have entrained to a given ZG, maximum and minimum staining intensity is expected at ZT 21 and ZT 9 respectively (Yoshii et al., 2009). Note that during sensory conflict ZT_L and ZT_T refer to the Zeitgeber time specified by light and temperature, respectively.

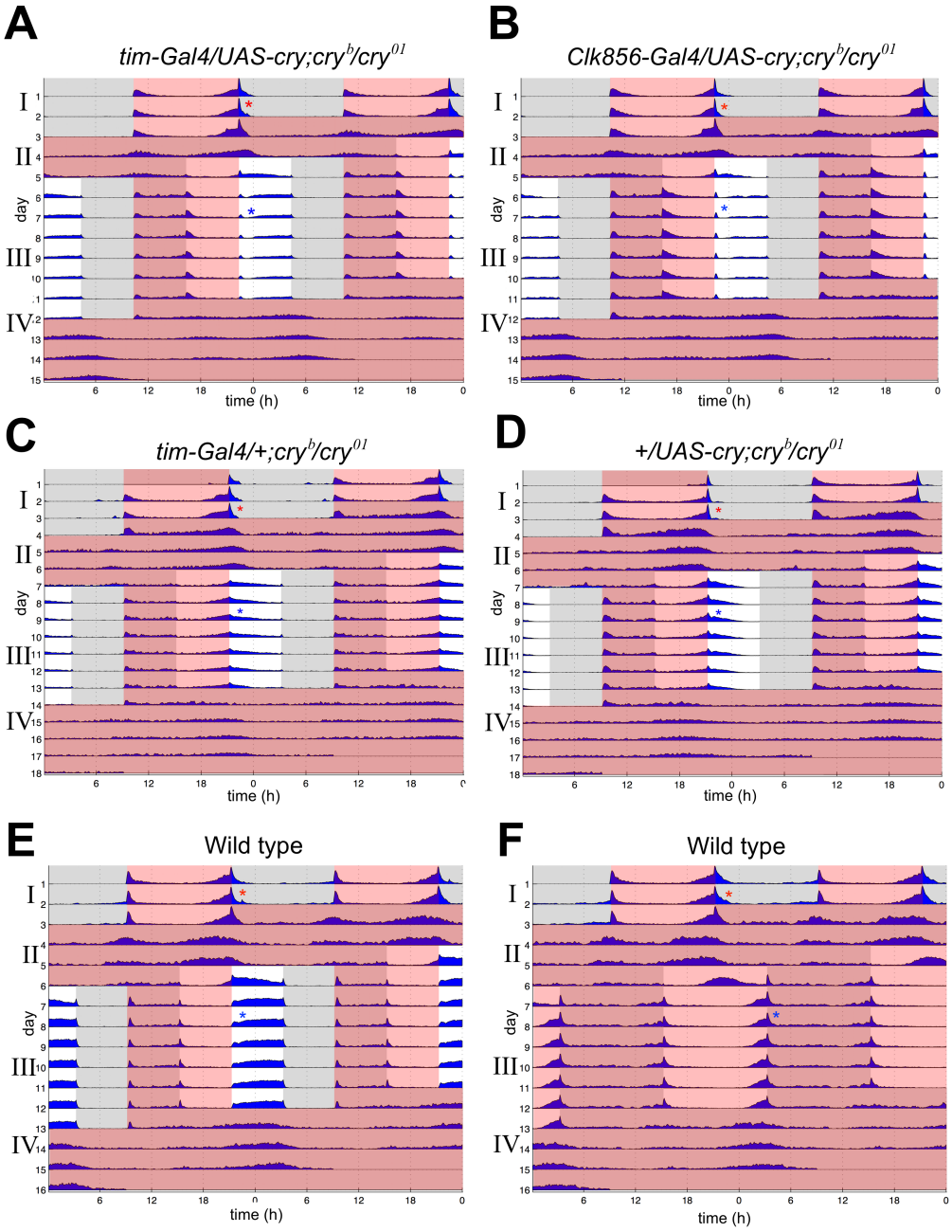

Figure 5.5: Circadian locomotor behaviour during sensory conflict cont. Average actograms of (A) *cry* mutant rescue using a *tim-Gal4* driver n = 32, (B) *cry* mutant rescue using a *Clk856-Gal4* driver n = 33, (C) *tim-Gal4/+;cry^b/cry^{01}* control (n = 12), (D) *+/UAS-cry;cry^b/cry^{01}* control (n = 21), (E) Wild type repeat (n = 38), and (F) wild type flies during LD jet lag experiment (Part III comprises 6 hr delay of LD relative to Part I, at constant 26°C) (n = 46). Environmental conditions outlined in Fig 5.3. Red filter shows warm phase; white background shows light phase. Red asterisk denotes representative evening behaviour in Part I; blue asterisk denotes representative pseudo-evening behaviour in Part III.

During aligned LD:TC (Part I), wild type flies showed the expected strong PER oscillations in all neuronal subgroups with a peak at ZT 21 and a trough at ZT 9 (Fig. 5.6 A left, 5.6 C Top). In cry^{02} mutants, PER cycled with the same phase, but with lower amplitude (Fig. 5.6B left), consistent with previous findings that light and temperature synergistically entrain molecular rhythms (Yoshii et al., 2009).

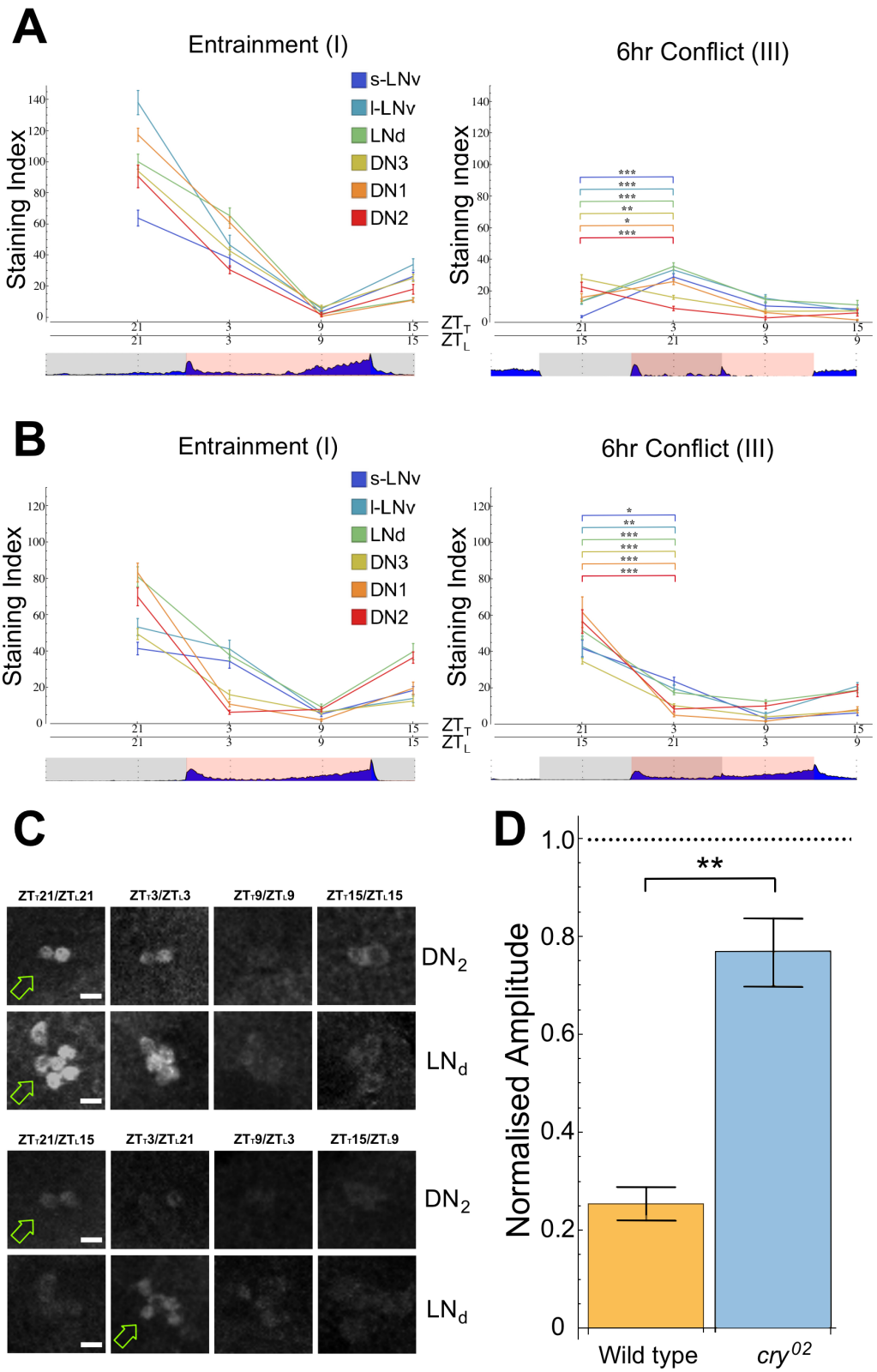
A
Entrainment (I)
6hr Conflict (III)
s-LNv
l-LNv
LNd
DN3
DN1
DN2
Staining Index
ZT_T
ZT_L
B
Entrainment (I)
6hr Conflict (III)
C
ZT21/ZT21
ZT3/ZT3
ZT9/ZT9
ZT15/ZT15
ZT21/ZT15
ZT3/ZT21
ZT9/ZT3
ZT15/ZT9
DN2
LNd
D
Normalised Amplitude
Wild type
cry02

Figure 5.6: Central clock molecular rhythms during sensory conflict. PER immunostaining of wild type (A) and cry^{02} (B) brains during 'Entrainment' (left: aligned LD:TC) and 6 hr conflict (right) conditions. One-way ANOVA reveals a significant effect of ZT on PER staining intensity during both aligned and misaligned conditions in both genotypes ($p < 1 \times 10^{-7}$ in all clock neuronal groups). During 6 hr conflict, t-test reveals significant differences between the first two time points plotted for all neuronal subgroups in wild type and cry^{02} mutants. Dissociation in peak staining between different neuronal groups occurred in wild type, but not in cry^{02} flies (see also Table 5.1). (C) PER staining in the DN2 and LN_d cell groups in wild type brains during entrainment (top) and 6 hr conflict (bottom) conditions. Scale bar = $5\mu m$. Green arrows mark maximum staining for each cell group. (D) Average amplitude of neuronal subgroup oscillations during sensory conflict (Part III) divided by that during aligned LD:TC conditions (Part I) in wild type and cry^{02}. A score of 1 denotes no change between conditions. All error bars represent SEM. ($p < 0.01$*, $p < 0.001$**, $p < 0.0001$***).

By contrast, during conflict, I observed a striking collapse in the amplitude of PER oscillations for all neuronal subgroups in wild type flies (Fig. 5.6 A right, 5.6 C Bottom, 5.6 D). Furthermore, inspecting the residual low-amplitude PER oscillations, there appeared to be a clear shift in the peak of the s-LN_v, l-LN_v, LN_d, and DN1 to ZT_L21, suggesting at least partial entrainment of these neurons to LD. In contrast, the CRY-negative DN2 and DN3 remained phase-locked to TC, displaying peak PER expression at ZT_T21. I did not notice any obvious phase heterogeneity within each neuronal subgroup (see for example the DN2 and LN_d in Fig. 5.6 C).

In cry^{02} mutants under conflict conditions, molecular rhythms remained comparable to Part I (Fig. 5.6 B right, 5.6 D). This echoes my behavioural findings, suggesting that the altered molecular rhythms observed in wild type flies result from the integration of conflicting inputs to the clock network, and that such conflicts can be avoided by weakening one of the input pathways, as in cry^{02} mutants.

5.3.3 Sustained effects of sensory conflict on the circadian clock

Considering the drastic effects of sensory conflict on behaviour and molecular clock oscillations, one might expect alterations to the underlying state of the circadian clock. This should manifest itself during constant conditions. I therefore analysed the consequences of sensory conflict (Part III) on the final free run section (part IV).

One-Way ANOVA: P-values:

Part I	s-LN$_v$	l-LN$_v$	LN$_d$	DN$_3$	DN$_1$	DN$_2$
Canton S	7.2×10^{-13}	6.6×10^{-27}	3.4×10^{-32}	5.8×10^{-37}	2.2×10^{-44}	2.9×10^{-21}
w;cry^{02}	2.1×10^{-13}	2.2×10^{-13}	1.7×10^{-22}	3.7×10^{-14}	6.9×10^{-29}	5.1×10^{-21}
Part III	**s-LN$_v$**	**l-LN$_v$**	**LN$_d$**	**DN$_3$**	**DN$_1$**	**DN$_2$**
Canton S	3.3×10^{-17}	5.0×10^{-16}	5.9×10^{-14}	1.5×10^{-13}	5.2×10^{-18}	3.4×10^{-7}
w;cry^{02}	3.8×10^{-17}	3.7×10^{-11}	3.4×10^{-19}	9.9×10^{-30}	5.9×10^{-21}	3.7×10^{-16}

Two-tailed T-Test Between ZT$_T$21 (ZT$_L$15) and ZT$_T$3 (ZT$_L$21) in Sensory Conflict: P-values:

	s-LN$_v$	l-LN$_v$	LN$_d$	DN$_3$	DN$_1$	DN$_2$
Canton S	1.4×10^{-14}	1.2×10^{-9}	5.8×10^{-12}	5.0×10^{-5}	1.6×10^{-3}	1.4×10^{-4}
w;cry^{02}	1.4×10^{-3}	1.8×10^{-4}	5.0×10^{-7}	3.0×10^{-16}	1.4×10^{-5}	1.5×10^{-6}

N numbers (brain hemispheres):

	Part I	s-LN$_v$	l-LN$_v$	LN$_d$	DN$_3$	DN$_1$	DN$_2$
ZT$_T$21	Canton S	19	19	20	21	21	21
	w;cry^{02}	11	11	12	11	12	12
ZT$_T$3	Canton S	25	23	26	24	25	21
	w;cry^{02}	18	18	18	18	18	16
ZT$_T$9	Canton S	18	18	18	18	18	13
	w;cry^{02}	23	23	23	24	24	24
ZT$_T$15	Canton S	21	22	23	24	23	21
	w;cry^{02}	12	12	12	12	12	12
	Part III	**s-LN$_v$**	**l-LN$_v$**	**LN$_d$**	**DN$_3$**	**DN$_1$**	**DN$_2$**
ZT$_T$21	Canton S	33	34	34	34	32	31
	w;cry^{02}	19	20	20	14	14	16
ZT$_T$3	Canton S	33	32	35	35	34	26
	w;cry^{02}	25	26	26	25	24	22
ZT$_T$9	Canton S	18	18	17	19	20	09
	w;cry^{02}	23	24	24	24	24	24
ZT$_T$15	Canton S	28	28	32	32	32	32
	w;cry^{02}	16	16	16	17	16	16

Table 5.1: Statistics to accompany antibody staining data presented in Fig. 5.6.

I compared overall rhythmicity and peak phase during the free run section, with that of control flies that had not experienced sensory conflict. These control flies were initially exposed to the identical in-phase LD:TC and free-running conditions (Part I and Part II), before being subjected to a 6 hr delayed LD cycle at constant 26°C (Part III, $\Delta t_{L,T}$ = 6 hr) and subsequent release into the final free run (Part IV, DD at 26°C) (Fig. 5.5 F).

While I did not observe any effects on overall rhythmicity or period length (Ta-

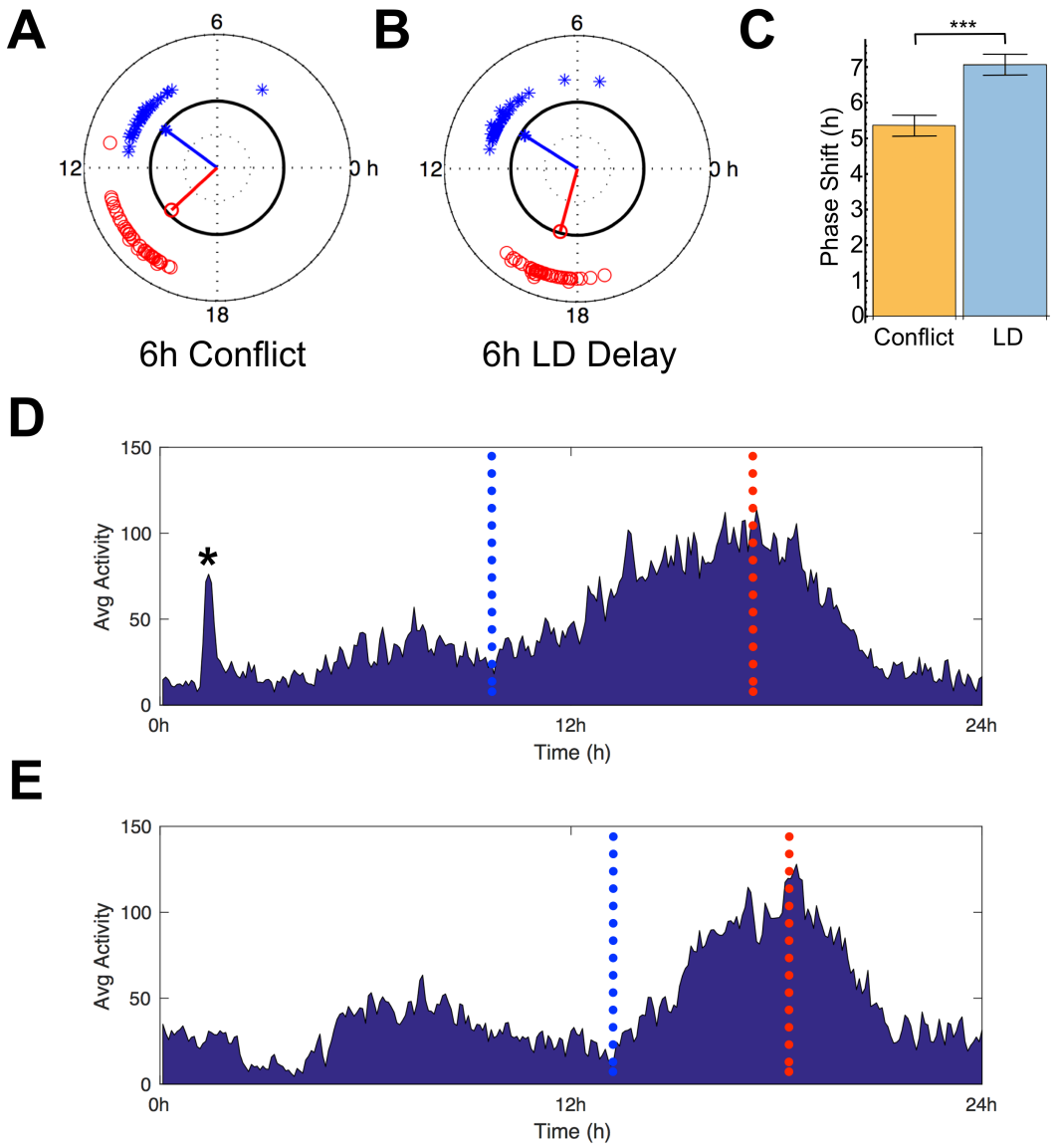

Figure 5.7: Sustained effects of sensory conflict on circadian clock phase. Comparison of the activity peaks during the free-running parts of the experiment (parts II and IV) preceding and following exposure to (A) 6 hr delayed sensory conflict (n = 38, phase difference = 5.3 hr, $p < 0.001$), or (B) 6 hr delayed LD cycle at constant 26°C (n = 46, phase difference = 7.1 hr, $p < 0.001$). Crosses show mean phase of each fly across the first 2 days of free run. Blue colour shows Part II; red colour shows Part IV. Circular statistics used are described in Batschelet (1981) (C) Bar chart showing magnitude of phase shift between Part II and IV in experimental groups (A) and (B). Error bars show standard deviation. $p < 1 \times 10^{-7}$. (D,E) Average locomotor behaviour in wild type flies during the first 2 days of free run following (D) 6 hr misalignment of LD:TC and (E) 6 hr LD shift in constant temperature. Average actograms covering the duration of experiment can be seen in Fig. 5.5E,F. Red dashed line denotes peak phase; blue dashed line denotes activity onset. Black asterisk denotes startle behaviour elicited by onset of temperature at the beginning of free run following conflicting conditions.

ble 5.2), suggesting similar clock function between the two groups, I did notice an advance of the evening activity peak in flies experiencing sensory conflict compared to those that were shifted with light at constant temperature (Fig. 5.7 A,B,C; Fig. 5.5 E,F). To quantify this apparent effect of the (un-shifted) temperature cue, I determined the magnitude of the phase difference between activity rhythms in free run Part II and Part IV for sensory conflict and control flies using circular phase analysis (Levine et al. 2002a; see Materials and Methods). As expected, both groups displayed almost identically phased activity peaks during Part II (2.4 hr and 2.2 hr before light and temperature onset in Part I, respectively). In contrast, in the free-run (Part IV) following sensory conflict, evening peak activity was delayed by 5.3 hr, while the peak of control flies was delayed by 7.1 hr (Fig. 5.5).

	τ(h)	RS	% Rhythmic	n
6 hr Conflict	24.6 ± 0.10	6.1 ± 0.20	100	38
LD Shift	24.3 ± 0.12	5.7 ± 0.19	95.7	44

Table 5.2: Quantification of free running activity rhythms in Part IV of the experimental regime. Free running period values (τ) and the corresponding strength of rhythms (RS) were calculated as described in Chapter 4. Flies with RS $\geq$ 1.5 were considered rhythmic.

Thus, exposure to conflicting ZGs appeared to diminish the degree of activity phase shift by almost 2 hr. This observation is consistent with theoretical considerations of the clock as coupled oscillatory subunits, which predict that the resulting equilibrium phase following conflicting input is some weighted average of the two inputs. This would act to reduce the degree of phase shift compared to synchronization with the 6 hr delayed LD alone.

5.3.4 Robustness of the Clock Network to Conflicting Inputs

A recent study by Yao and Shafer (2014) suggests that the *Drosophila* central clock network is resilient to period discrepancies between neuronal subgroups, such as PDF-negative and PDF-positive neurons. Indeed, it was shown that coherent activity rhythms could still be generated, provided the period length mismatch between the cell groups was less than 2.5 hours. Having shown that 6 hr-misaligned LD:TC generates novel P activity patterns (Fig. 5.3 B; Fig. 5.5 E) associated with a severe

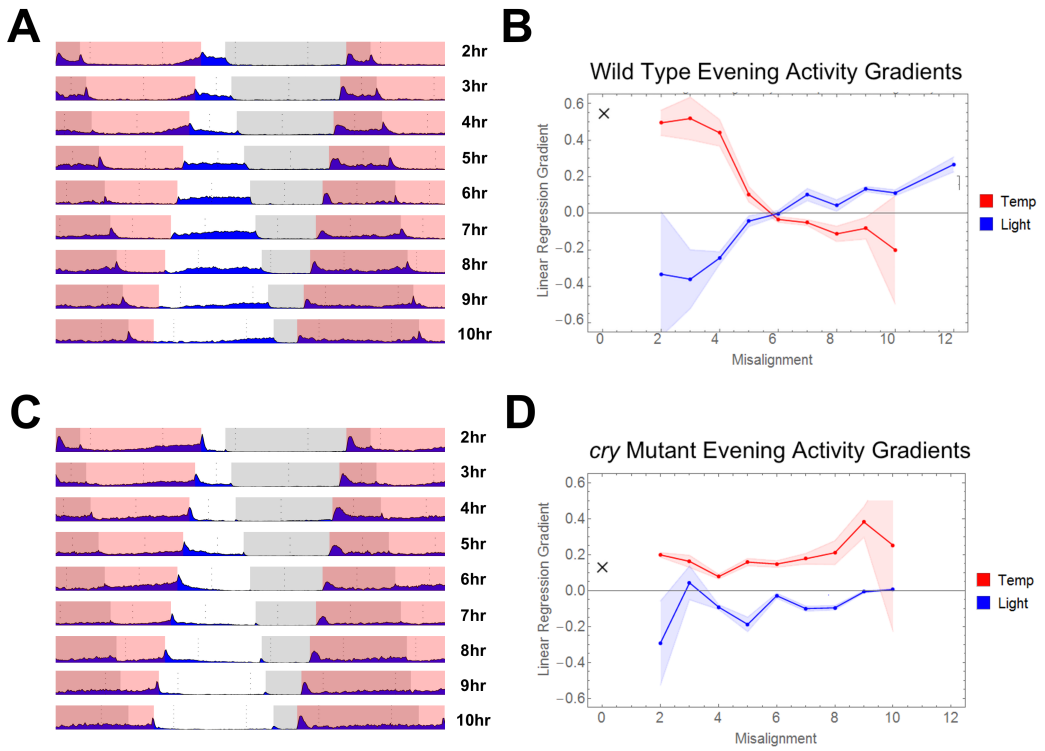

Figure 5.8: Behavioural responses to different light and temperature phase relationships. Varying degrees of LD:TC misalignments in wild type (A, B) and cry^{02} flies (C, D) ($45 \leq n \leq 65$). (A, C) Representative days of locomotor behaviour taken from average actograms during conflict conditions after activity rhythms had stabilised (Part III, day 5-6). (C, D) Gradient (*b* parameter from page 70) of linear regression fit to the period of activity preceding light and temperature transition to 'off'. Shaded regions denote 95% confidence intervals. Black cross indicates gradient of evening activity during corresponding in-phase condition. (Refer back to page 70 and Fig. 5.2).

collapse of PER oscillations in all clock cell groups, and phase differences between light- and temperature-sensitive clock subgroups (Fig. 5.6 A,D), I went on to explore the consequences of other degrees of sensory conflict for circadian locomotor behaviour. Adapting Part III of my experimental regime, I conducted a systematic behavioural analysis investigating the effect of varying the magnitude of the LD delay.

When LD:TC misalignment was less than 4 hr, wild type flies displayed anticipatory behaviour and peak activity at the end of thermophase, thus appearing to primarily follow the temperature signal (Fig. 5.8 A). However, activity did persist after temperature offset into the lights-on phase, suggesting some effect of light

on circadian locomotor behaviour (reminiscent of that observed in cry^{02} mutants during 6 hr conflict). Fully-fledged P behaviour emerged at 5-7 hr misalignments, with an absence of conventional entrainment to either signal at these degrees of sensory conflict. As the disparity between LD:TC exceeded 7 hr-misalignment, P behaviour gradually decayed, with a discernible evening peak of activity observed at light offset during 10 hr-misaligned conditions. These results go some way towards explaining previous observations made in complete antiphasic (i.e. $\Delta t_{L,T}$ = 12 hr) light and temperature (Yoshii et al., 2010) − only during very large sensory conflicts is light the dominant ZG.

In an effort to quantify my observations, I developed a mathematical framework to analyse the degree of fruit fly locomotor entrainment to a ZG (see Materials & Methods on page 70). This involved assessing the linear gradient of the anticipatory increase in locomotor behaviour that arises towards the offset of ZGs during entrainment.

When LD and TC are synchronized, this linear gradient was positive, consistent with the presence of evening anticipation. During misaligned conditions, the two separate gradients associated with light and temperature offset were used to gauge the disruption caused. The progressive change in these gradients for both temperature and light, as the degree of misalignment is varied, is shown in Fig. 5.8 B. P behaviour occurs when both gradients approach zero (see zero-crossing at 6 hr misalignment in Fig. 5.8 B). Entrainment to TC for smaller misalignments, and to LD for the largest misalignments, is also evident from the plot, as linear gradients of locomotor activity are more positive at these times respectively.

In contrast to wild type flies, the activity rhythms of cry^{02} mutants remain largely entrained to TC, independent of the magnitude of the sensory conflict (Fig. 5.8 C). This unwavering temperature preference is again illustrated numerically by the fact that temperature evening gradients remain more positive than light evening gradients for all LD:TC misalignments (Fig. 5.8 D).

5.4 Discussion and Future Work

Circadian research in *Drosophila melanogaster* has traditionally treated light and temperature separately. However, clock networks evolved to orchestrate behaviour within multisensory environments. Recent studies suggest the existence of multiple independent oscillatory subunits within the fly central clock, each capable of driving activity patterns (Yao and Shafer, 2014). Such distributed architectures tend to exhibit cooperation and/or competition. In this chapter of my thesis, I have performed a systematic and quantitative exploration of the behavioural, and molecular effects of conflicting (light/temperature) entrainment regimes on the circadian system. My paradigm offers a novel route to decompose the circadian network, and my findings demonstrate that sensory conflict can, under specific conditions, cause dramatic disruptions to clock output, which have not been reported before.

Although light does indeed dominate temperature for maximal misalignments, I have shown that smaller delays of LD relative to TC lead to evening activity rhythms in wild type flies that are predominantly entrained to the temperature cue. These observations are in line with previous reports of temperature also being the critical parameter for morning activity onset in natural conditions (Vanin et al., 2012). Accordingly, my findings indicate a greater biological relevance for temperature effects on daily behavioural rhythms than previously appreciated.

Furthermore, with larger delays of 5-7 hours, typical evening peaks of activity broke down, giving way to an abnormal locomotor pattern, which I here refer to as Plateau (P) behaviour. This P behaviour is associated with a drastic reduction in the amplitude of molecular rhythms in the central clock, as well as dissociation between clock neuronal groups. Importantly, 6 hr sensory conflict between light and temperature also reduced the degree of phase shift compared to that induced by 6 hr delay of light alone. This demonstrates that sensory conflict still acts to entrain the circadian oscillator (Fig. 5.7), while it does not bring about any long-term changes to clock function, as shown by the rhythm strength and periodicity of the locomotor behaviour after exposure to sensory conflict as compared with light alone (Table 5.2).

It was only during even larger misalignments of 8-10 hr that a restoration of more typical evening activity peaks was seen, and a reversal of cue preference back to the light signal (cf Yoshii et al. 2010). Together, these results emphasize the context-dependent nature of ZG dominance. The *Drosophila* circadian system, it appears, is able to generate wild type-like behavioural rhythms only for a limited range of light-temperature phase relationships, i.e. either very small or very large misalignments; intermediate conflicts, however, are not easily accommodated by the central clock network.

Throughout my investigation, I have maintained a phase-agnostic approach to my experimental interpretations. It remains unclear how the phase of environmental oscillatory signals translates to circadian phase extracted by the clock. Indeed, temperature typically lags behind light under natural conditions (Boothroyd et al., 2007; Vanin et al., 2012), suggesting that $\Delta t_{L,T} = 0$ hr might not necessarily represent 'in-phase' signals as far as the clock is concerned. Pending deeper understanding, we must only treat phase relationships between light and temperature in a relativistic manner. Thus, the coincidence of photo- and thermo-phases should be thought of as an arbitrary reference point (admittedly, one that has been used frequently in the field).

From a mechanistic viewpoint, my molecular data reveal a striking effect of sensory conflict, as 6 hr LD:TC misalignments lead to a drastic reduction in the amplitude of molecular rhythms in all clock neurons. The phase of the remaining low-amplitude oscillations appears largely consistent with that reported previously (Yoshii et al., 2010), revealing a temperature preference of the *cry*-negative DN2 in wild type flies.

Curiously, residual PER rhythms in the DN3 also appeared to align with TC during sensory conflict. This finding, which might well be linked to my particular experimental conditions, has not been reported previously – in 12 hr conflict conditions, PER rhythms in DN3 were reported to entrain to light (Yoshii et al., 2010). My results do, however, resemble the TIM cycling reported previously during sensory conflict (Miyasako et al., 2007), suggesting a temperature-sensitive property

of the DN3.

5.5 Summary

Robustness towards a range of variable, and potentially conflicting, inputs is a beneficial property for any sensory network. In this chapter, I have shown that phase discrepancies between clock neurons can result from sensory conflict, and that in these conditions, the fly clock resists some, but embraces other, misalignments. In this way, I hope to have addressed objectives 1 and 2 outlined in Chapter 3.

Network robustness offers obvious advantages in itself, but possible benefits extend beyond this. Resilience might also imply plasticity, allowing different clock cell groups to exhibit autonomy under different conditions, truly optimizing behaviour for particular environmental features. Moreover, in nature, the phase relationship between light and temperature might also provide valuable 'circannual' information to the network, owing to seasonal changes between these two cues.

Building on previous studies, I focused on the interplay between light and temperature in *Drosophila*. It is important to note, however, that the impact of my findings are not necessarily restricted to these two sensory entrainment pathways, nor indeed are they restricted to the fly. Links between human circadian clock function (and dysfunction) and mental disorders have been made repeatedly, but the directions of the underlying causalities remain unclear (Roenneberg and Merrow, 2016). Most intriguing in this regard is the suggestion that the associations between psychiatric pathologies and the clock partly involve behavioural habits, which alter an individuals exposure to different ZGs (Adan et al., 2012). A more thorough study of multisensory processing in the circadian system, and possible conflicts that can arise therein, therefore stands not only to increase our understanding of the computation of time, but also to enable novel approaches in the treatment, and prevention, of mental disorders.

As I outlined in Chapter 1, page 33, multiple other sensory cues are capable of entraining the fruit fly clock (Fujii et al., 2007; Levine et al., 2002c; Simoni et al., 2014). I further believe it likely that more input signals may be revealed with

increased research in the field. The principles of my results, therefore, should bear increased relevance moving forward – particularly regarding context-dependent ZG dominance, and the circadian disruption that can occur at certain environmental phase relationships.

Throughout this investigation, I was guided by the theoretical considerations outlined in Section 1.4.2 and Section 5.1. These proved extremely useful in specifying the form of environmental regime that might sufficiently challenge the circadian network. Moreover, Bayesian reasoning provides vital explanatory power when considering my results in the context of previous studies. This highlights the benefits of a robust dialogue between theory and experimentation in the field of chronobiology.

Chapter 6

Light Dominates Peripheral Circadian Oscillations in *Drosophila* During Sensory Conflict

The results presented in this chapter have been published in the *Journal of Biological Rhythms*
(see Harper et al. 2017)

6.1 Introduction

In the previous chapter, I discussed results that highlight the exquisite sensitivity, and thus vulnerability, of the circadian system in *Drosophila* to environmental phase relationships; tolerating only certain degrees of Zeitgeber mismatch. We saw that a 6 hr misalignment between LD and TC leads to reduced-amplitude PER oscillations in central clock neurons, dissociation between different light- and temperature-sensitive cell groups, and an associated loss of evening anticipation behaviour (Harper et al., 2016). However, as outlined in Chapter 1, the circadian system in *Drosophila* extends beyond the confines of the fly head, to peripheral tissues around the body (Giebultowicz and Hege, 1997; Giebultowicz et al., 2001; Ito et al., 2008; Plautz, 1997).

The entrainment of peripheral clocks to single ZGs has been studied. For instance, many respond directly to both light and temperature, entraining PER

rhythms to LD and TC when isolated *in vitro* (Glaser and Stanewsky, 2005; Ivanchenko et al., 2001; Plautz, 1997). As in the central clock, light sensitivity of peripheral clocks appears to act via CRY, which mediates the light-dependent degradation of TIM (Ivanchenko et al., 2001; Stanewsky et al., 1998). Unlike the central clock, however, peripheral CRY may serve a dual function as a core clock component (Collins et al., 2006; Krishnan et al., 2001; Levine et al., 2002b) (discussed in Chapter 1, page 29). This latter role of CRY resembles that which is observed in mammalian systems (Okamura, 1999; van der Horst et al., 1999).

While unimodal entrainment of peripheral clocks has been demonstrated, the responses of these oscillators in multisensory environments remains unknown. Thus, building on the results presented in Chapter 5, I asked how peripheral circadian networks respond to sensory conflicts between light and temperature. To achieve this, I used a well-established *period-luciferase* fusion gene (*XLG-luc*) reporting PER expression in peripheral clocks (see Materials & Methods, Section 6.2; Veleri et al. 2003). Furthermore, given that peripheral clocks form part of a wider circadian network throughout the fly, I chose to investigate responses *in vivo* rather than purely in isolated tissues.

In this chapter, I will show that, unlike the central clock network, peripheral clock oscillations do not collapse under sensory conflict. Instead, light dominates peripheral clock entrainment and this light dominance depends on the circadian photoreceptor, CRY.

6.2 Materials & Methods

6.2.1 Fly Strains

Flies were reared under 12 hr:12 hr LD cycles on *Drosophila* medium (0.8% agar, 2.2% sugar-beet syrup, 8.0% malt extract, 1.8% yeast, 1.0% soy flour, 8.0% corn flour, and 0.3% hydroxybenzoic acid) at 25°C and approximately 60% humidity. The following fly stocks were used: *XLG-luc1-1* (Veleri et al., 2003), tim^{01}*;XLG-luc1-1* (Glaser and Stanewsky, 2005), and *XLG-luc1-1* cry^{b} double mutants generated by meiotic recombination between *XLG-luc1-1* and cry^{b} bearing chromo-

somes. Only male flies between 3 and 6 days old were used in experiments.

6.2.2 Activity Monitoring

Locomotor activity rhythms were recorded automatically using the *Drosophila* Activity Monitoring (DAM) system (*Trikinetics*, Waltham, MA) as previously described in Chapters 1 and 4. Analysis of locomotor data was conducted as described in Chapter 4.

6.2.3 Zeitgebers

Field studies show that temperature rises continuously (and almost linearly) throughout the day and falls in a similar fashion during the night (Vanin et al., 2012). However, apart from some recent exceptions (e.g. Yoshii et al. 2010), it has been traditional to use sharp, square-wave-like transitions between cold and warm conditions to study temperature-dependent circadian entrainment. My preliminary luciferase assay experiments showed that such square-wave-like temperature cycles gave rise to sharp peaks in bioluminescence readings at the transitions between cold and warm (see Fig. 6.1 for examples). The fact that these peaks were also present in clock-null mutant flies (tim^{01}; Fig. 6.1 bottom trace) suggests they are extrinsic to the circadian oscillator, possibly reflecting an altered metabolic activity during sharp temperature transitions (Sehadova et al., 2009).

At a conceptual level, I draw a comparison with masking, which commonly describes direct stimulus-evoked (i.e. clock independent) changes in locomotor behaviour at points of sharp environmental transition (see Table. 1.2). By analogy, the clock-independent bioluminescence responses could be described as a form of 'molecular masking', which might well be linked to the behavioural responses. To mitigate these masking effects, I decided to use more naturalistic, 'ramped' temperature cycles (Vanin et al., 2012), which made the transitional bioluminescence responses disappear (as is also the case with behavioural masking) (Fig. 6.1).

12:12 light-dark cycles were generated through square wave transition between 2500 and 0 lux respectively. 12:12 temperature cycles were achieved through gradual transitions between 26°C and 16°C occurring over 9.5 hours (see Fig. 6.2 A,B).

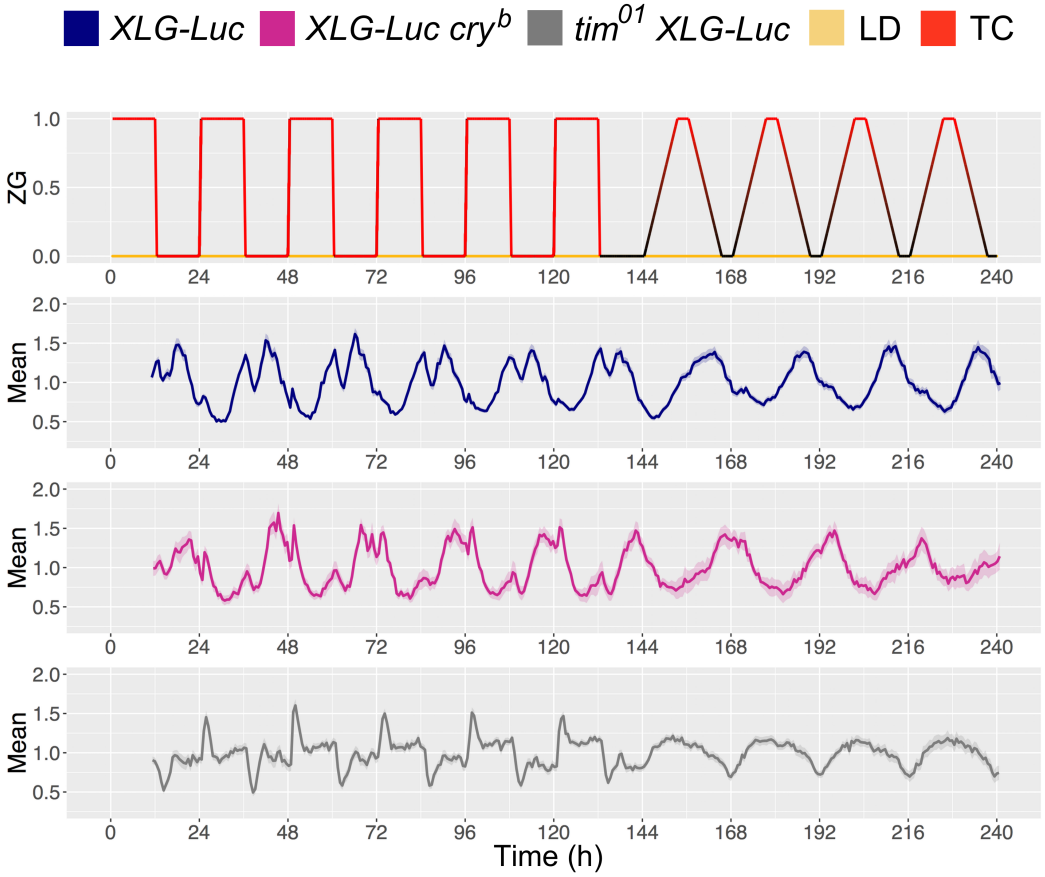

Figure 6.1: Bioluminescence recordings in square-wave and ramping temperature cycles. Dark blue, magenta and grey lines show mean average of *XLG-luc*, *XLG-luc1-1 cry^b* and *XLG-luc tim^{01}* flies, respectively. Shaded regions show SEM. Experimental regime in which 5 days of square-wave-like TC were followed by 4 days of ramped TC (16:26°C and DD in both cases, see upper panel). (n = 30, n = 16 and n = 24 for wild type, *cry* mutant, and *tim* null flies respectively). Note the marked changes in bioluminescence responses between the two entrainment regimes, which can be seen across all three experimental lines.

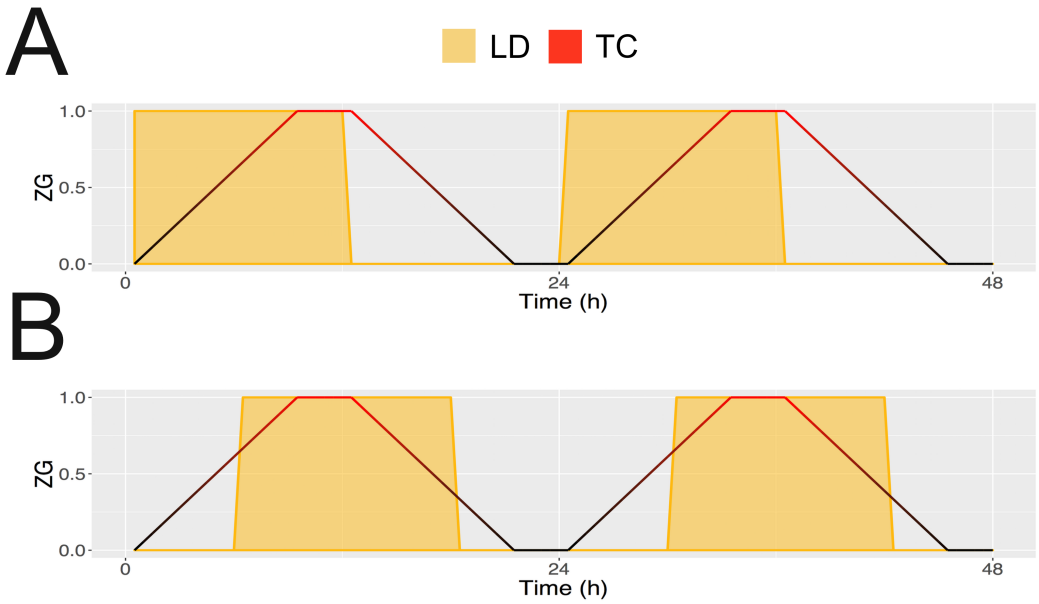

Figure 6.2: Environmental LD:TC regimes. Light and naturalistic temperature cycles are either aligned (A) or misaligned by 6 hr (B).

ZT 0 denotes the beginning of an increase from 16°C to 26°C over 9.5 hours, and ZT 12 denotes the start of a corresponding decrease from 26°C to 16°C.

As in Chapter 5, cue misalignment is quantified as the absolute distance, in hours (delta time, or Δt), between the onset/offset of two cyclic 12:12 signals. For example, $\Delta t_{L,T}$ = 3 hr denotes that light onset/offset occurs 3 hr after the beginning of temperature rise/fall.

Environmental conditions were recorded with an environmental monitor placed inside the incubator. These were checked to validate scheduled conditions.

6.2.4 Bioluminescence Imaging

The images taken in Fig. 6.4 were acquired using an LV200 bioluminescence imaging system (*Olympus*). Flies were kept in 12:12 LD at 25°C (Fig. 6.4 A,B) or 12:12 naturalistic TC in DD (Fig. 6.4 C,D) for at least 3 days in vials containing a 1% agar and 5% sucrose food supplemented with 15mM luciferin. Individuals were then anaesthetized using ether and dry mounted under a UPLSAPO 20x Apochromat objective. Light microscopy was used to obtain a reference image of the fly. The lights were then turned off and a CCD camera (*Hamamatsu* ImageEM X2) was used to capture the bioluminescence signal over a 5 min exposure period.

All images were processed and quantified using the same settings in FIJI (Schindelin et al., 2012). Luminescence was assessed by measuring the mean pixel intensity of either the abdomen or the head body sections. From this mean value, a background pixel intensity was subtracted to produce the final quantifications shown in Fig. 6.4 B,D.

6.2.5 Bioluminescence Rhythms Acquisition

Bioluminescence assays were performed as previously described (Glaser and Stanewsky, 2005). 3-6 day-old male flies were placed in alternate wells of a 96-microtiter plate (48 flies per plate). Each well contained 100μl luciferin medium (1% agar, 5% sucrose, and 15mM luciferin (*Biosynth*, Switzerland)). Using a Packard TopCount Multiplate Scintillation Counter (*Perkin-Elmer*) that was placed in a light- and temperature-controllable incubator (*Percival*), the photon count per second (CPS) emitted from each well was measured every 60 minutes for 15 seconds per well, unless otherwise stated in the text. Flies were exposed to the experimental regime as specified at relevant points in Results.

If at any point in the experiment individual well emissions dropped to background levels (approximately 50-100 CPS), flies were regarded as dead, and their data were excluded from the analysis from this point onwards.

6.2.6 Bioluminescence Data Analysis

Bioluminescence rhythms were analysed as described in Chapter 4 using the custom-built *Luceo* toolbox developed in *MATLAB*.

6.3 Results

6.3.1 Spatial Expression Pattern of the *XLG-luc* Transgene

I chose to study bioluminescence changes using the *XLG-luc* transgene. *XLG-luc* contains the endogenous *period* promoter, driving the expression of a *period-luciferase* fusion gene. The XLG-LUC protein is expressed in most, if not all, per-expressing cells (Veleri et al., 2003), including neurons in the fly brain (Veleri et al., 2003), as well as those in the periphery (Glaser and Stanewsky, 2005). However, the

construct is widely used as a peripheral clock reporter, since signals from the central clock are expected to be overwhelmed by those from the peripheral *per*-expressing cells (Glaser and Stanewsky, 2005; Sehadova et al., 2009; Veleri et al., 2003). Evidence for this comes from the fact that rhythmic photon counts in *XLG-luc* flies are by a factor of ~250-times greater than those observed for an *8.0-luc* transgenic line, in which *luciferase* expression is restricted to central clock cells (Veleri et al., 2003). Nevertheless, exactly which *per*-expressing tissues predominantly contribute to the bioluminescence signal of *XLG-luc* flies is not known.

To visualize the source of the *luciferase* reporter signal that is measured in the time series assays, I decided to perform whole-animal bioluminescence imaging on flies that had been reared in circadian light and temperature conditions. Due to the presence of superior imaging equipment in the lab of my supervisor, Ralf Stanewsky (Münster, Germany), the raw data presented in Fig. 6.4 were obtained by Maite Ogueta, a post-doctoral researcher in the lab. I personally designed the imaging protocol, analysed the data, and provided technical guidance for the experiments.

To identify appropriate times for the imaging of maximum and minimum PER expression in *XLG-Luc1-1* and *XLG-luc1-1 cry^b* flies during both LD and TC separately, I first performed bioluminescence assays for these unimodal conditions (Fig. 6.3). These results show that peak and trough PER expression in LD and TC alone occurs between ZT 19-23 and ZT 7-11 respectively. This is consistent with previous findings (Glaser and Stanewsky, 2005). Also clear from the bioluminescence traces is that strong peripheral PER rhythms are seen in wild type flies during both LD and TC, yet *cry* mutants show peripheral rhythms only in TC (see also Fig. 6.4 A-D).

XLG-luc transgenic flies in wild type and *cry* mutant background were accordingly imaged between ZT 19-23 and ZT 7-11 in LD and naturalistic TC, when PER expression is expected to be at peak and trough levels respectively (Fig. 6.3; Glaser and Stanewsky 2005). Bioluminescence levels varied markedly between time points for wild type flies in both LD and TC. However, a similar change in *cry* mutants was only observed during TC, consistent with a drastically weakened light input pathway in flies lacking functional CRY (Fig. 6.4 A-D; Glaser and Stanewsky 2005;

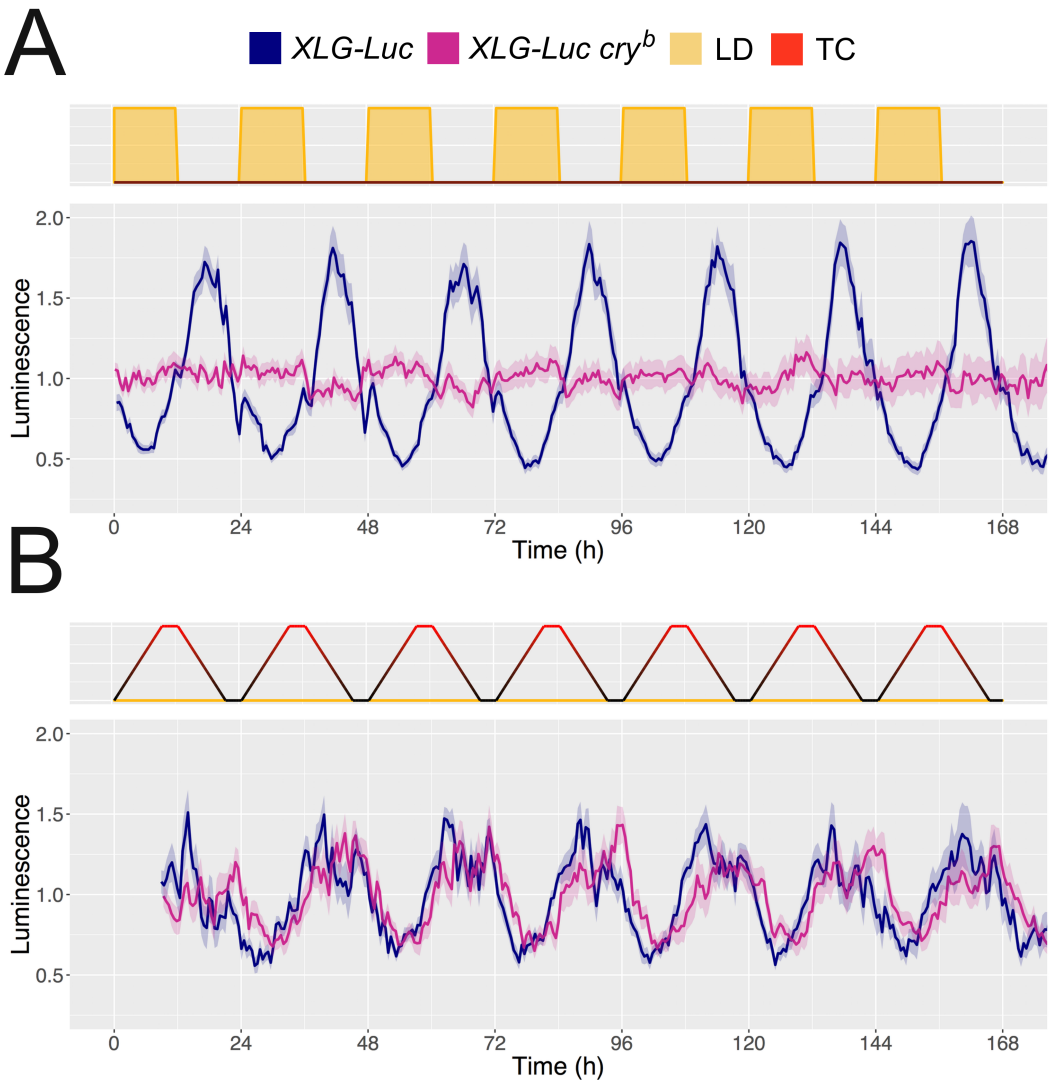

Figure 6.3: Bioluminescence assays in LD and TC alone. (A) Experimental regime in which environmental conditions followed 7 days of LD at 26°C. (n = 18 and n = 22 for wild type and *cry* mutant flies respectively). Median peak phase of *XLG-luc* flies was at ZT 18.21; no discernible peak phase for *XLG-luc cry^b*. (B) Experimental regime in which environmental conditions followed 7 days of ramped TC (16:26°C) in DD (n = 17 and n = 22 for wild type and *cry* mutants respectively). Median peak phase of *XLG-luc* and *XLG-luc cry^b* was at ZT 16.65 and ZT 18.83 respectively. Bioluminescence readings acquired at 30 min resolution.

Stanewsky et al. 1998). This finding is again confirmed in the bioluminescence time series assays shown in Fig. 6.3.

In both genotypes, an overwhelming majority of signal was seen emanating from the abdomen and eyes (Fig. 6.4 A,C), consistent with that observed in other *per-luc* transgenics (Stanewsky et al., 1997a). We can therefore conclude that *in vivo* assays monitoring bioluminescence changes in *XLG-luc* reporter lines will be dominated by signals from these particular peripheral clock components.

6.3.2 Sensory Conflict Generates P-like Behaviour under Naturalistic Temperature Cycles

In Chapter 5, we saw that a 6 hr misalignment between LD and (square wave-like) TC can disrupt normal circadian locomotor patterns in Canton S wild type flies, characterized by a complete loss of evening anticipation. In place of this, flies display Plateau (P) behaviour, which features a period of sustained high activity, bordered by relative inactivity (Fig. 5.3 B; Fig. 5.4; Fig. 5.5 E; Harper et al. 2016). This P behaviour is also associated with a breakdown of molecular oscillations in central clock neurons (Fig. 5.6 A).

It was therefore important to examine whether P behaviour is also present in *XLG-luc* transgenic flies during misaligned LD:TC for temperature cycles that are naturalistic. An environmental regime was administered comprising aligned LD:TC ($\Delta t_{L,T}$ = 0 hr) for 3 days, followed by incremental 2 hr delays of LD ($\Delta t_{L,T}$ = 2 hr, $\Delta t_{L,T}$ = 4 hr) to generate what was ultimately a 6 hr misalignment ($\Delta t_{L,T}$ = 6 hr) for 3 days.

During aligned LD:TC, *XLG-luc* flies displayed a characteristic bimodal profile of activity, with evening anticipation and a peak coinciding with the end of photo/thermo-phase (Fig. 6.5 C). During, 6 hr misaligned LD:TC, *XLG-luc* locomotor activity displayed P behaviour as observed previously using rectangular TC (Fig. 6.5 B,D; Harper et al. 2016) − forming a broad, flat bout of activity between lights-off and the beginning of falling temperature.

As expected, there were some differences in the behavioural pattern between misaligned LD:TC using naturalistic TC, compared to that using rectangular TC.

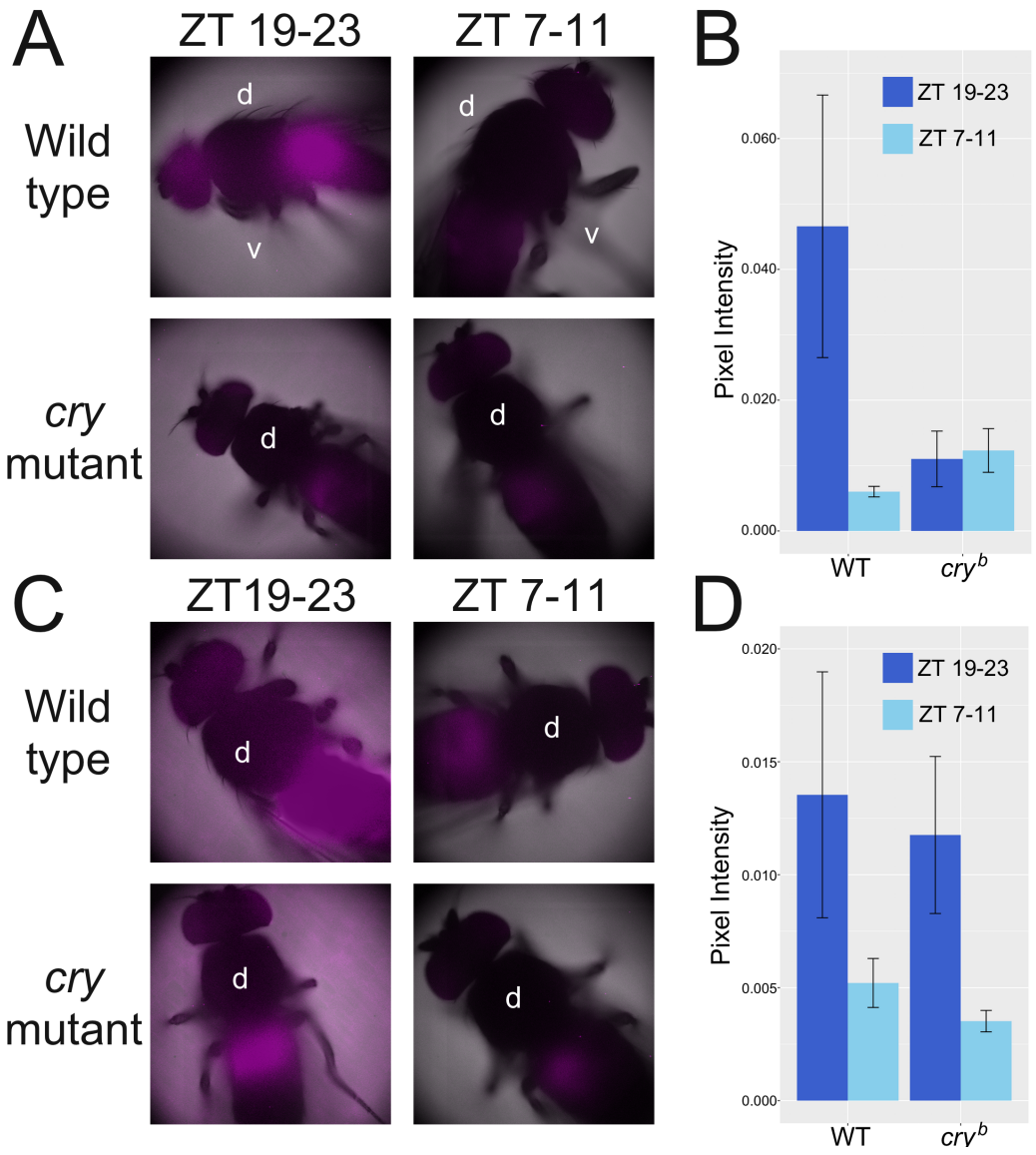

Figure 6.4: Bioluminescence imaging of *XLG-luc* transgenic flies. Pseudo-colour image of *XLG-luc* bioluminescence superimposed on a bright-field (black and white) reference image. Male *XLG-luc* and *XLG-luc* cry^b flies were fed on luciferin fortified food (15 mM) for 3 days in LD (A,B) or 16:26°C ramping TC (C,D) conditions, and imaged for 10 min with the LV200 Bioluminescence imaging system (*Olympus*). Images were taken between ZT 19 to ZT 23, and between ZT 7 to ZT 11, reflecting PER peak and trough expression levels, respectively. Panels B and D show quantifications of abdominal bioluminescence intensities for both LD (B) and TC (D) entrainment regime at the different circadian times indicated. The orientation of the fly is specified by the label 'd' (dorsal) and 'v' (ventral). Note the main signal sources in the abdomen and eyes. B) n = 7/n = 8, and n = 4/n = 10, for peak/trough expression in wild type and *cry* mutant flies respectively. D) n = 10/n = 8, and n = 7/n = 5, for peak/trough expression in wild type and *cry* mutant flies respectively. Error bars indicate SEM.

XLG-luc flies did not exhibit the sharp activity increases at the beginning of the warm phase, nor did they show a rapid drop in activity after lights-off (compare Fig. 6.5 B and D). Instead, *XLG-luc* flies showed a smoother activity increase during the rising temperature phase, punctuated by the lights-on transition, and a similarly smooth activity decrease during the falling temperature phase after lights-off (Fig. 6.5 D). It seems reasonable to attribute these minor differences to the fact that naturalistic temperature cycles were applied (cf. Yoshii et al. 2009).

Importantly, P behaviour was observed during misaligned naturalistic TC and LD conditions, demonstrating that the clock's behavioural output during sensory conflict is broadly similar between rectangular and naturalistic entrainment conditions, and can be observed in the transgenic *XLG-luc* flies used in the *luciferase* assays.

6.3.3 Light Dominates Peripheral Clock Entrainment in Wild Type Flies

Having confirmed the presence of P behaviour in *XLG-luc* flies, I next sought to investigate the response of the peripheral clock system during conflicting 6 hr phase-shifted LD:TC using naturalistic temperature cycles. Stability of peripheral endogenous rhythms was first assessed in free running conditions (DD:26°C) after entrainment to aligned and misaligned LD:TC (Fig. 6.6 A,B).

To investigate potentially richer forms of conflict, further experiments were conducted in which the period of free running conditions was followed by either aligned or misaligned LD:TC to contrast that experienced in the first part of the experiment (Fig. 6.6 C; Appendix Fig. A.1 A,B). All experiments were performed in *XLG-luc* flies in both wild type and cry^b mutant genetic backgrounds.

During aligned conditions, both wild type and cry^b flies displayed rhythms of bioluminescence, peaking during the night, as defined by both light and temperature (Fig. 6.6 A,C; Fig. 6.7 left; Fig. A.1 A,B). This observation agrees with previously reported PER protein oscillations in wild type flies during unimodal LD and TC entrainment (Glaser and Stanewsky, 2005), and also agrees with my own observations (Fig. 6.3).

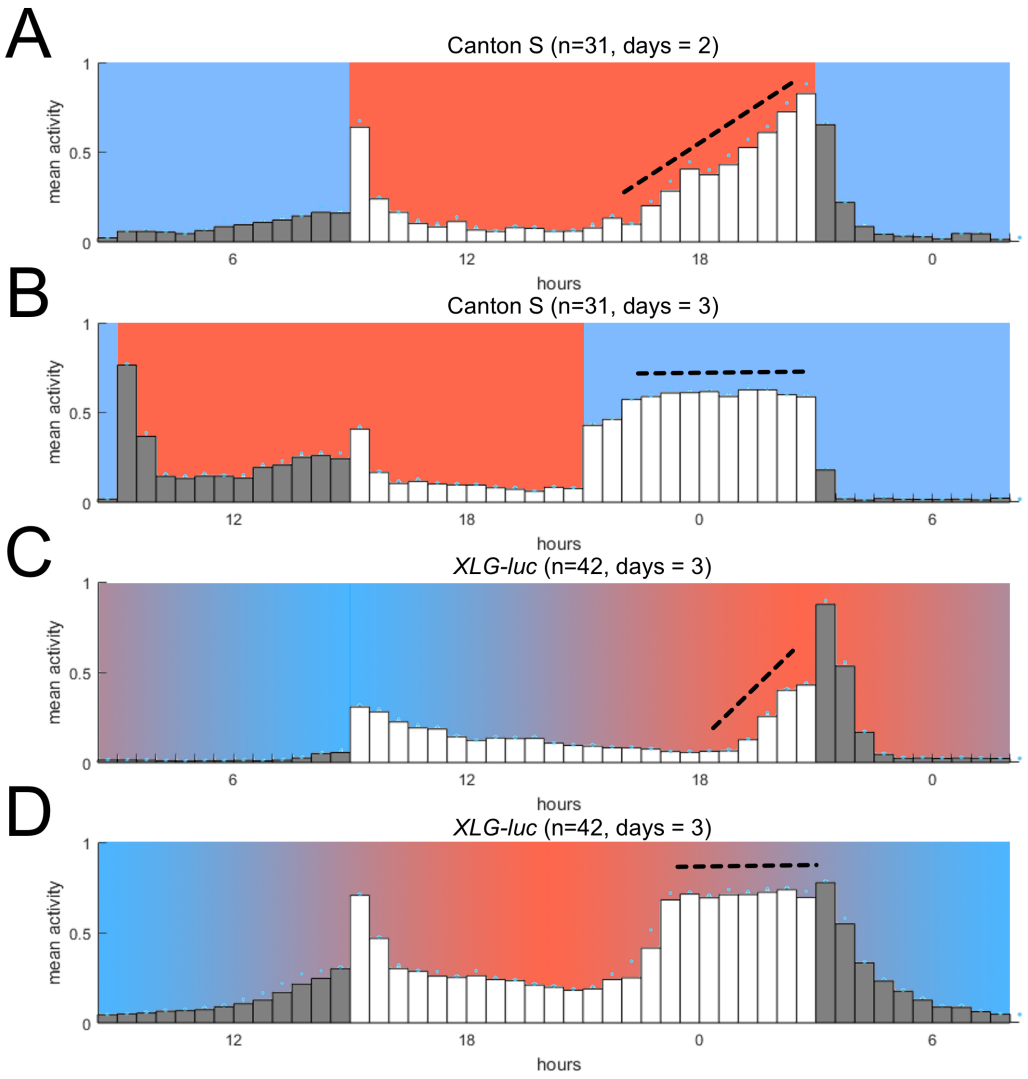

Figure 6.5: Locomotor behaviour during sensory conflict using square-wave and naturalistic TC. (A) Average locomotor activity in aligned LD:TC (lights-on coincides with temperature-on, n = 31). (B) Average locomotor activity in 6 hr misaligned LD:TC (lights-on delayed by 6 hours relative to temperature-on, n = 31). (C,D) New data in which a naturalistic temperature regime was used to generate sensory conflict. (C) Average locomotor activity in aligned LD:TC (lights-on coincides with the start of temperature-rise, n = 42). (D) Average locomotor activity in 6 hr misaligned LD:TC (LD delayed by 6 hours relative to temperature-rise, n = 42). Locomotor data for each plot was rescaled in the range of 0 and 1 to facilitate profile comparisons. Mean raw activity across days (total beam breaks/5 min/fly) was 13.3 (A), 11.8 (B), 10.4 (C), and 17.9 (D). Black dashed lines highlight presence/absence of evening anticipation. (Data in A,B adapted from Harper et al. (2016) in which a square-wave temperature regime was used to generate sensory conflict).

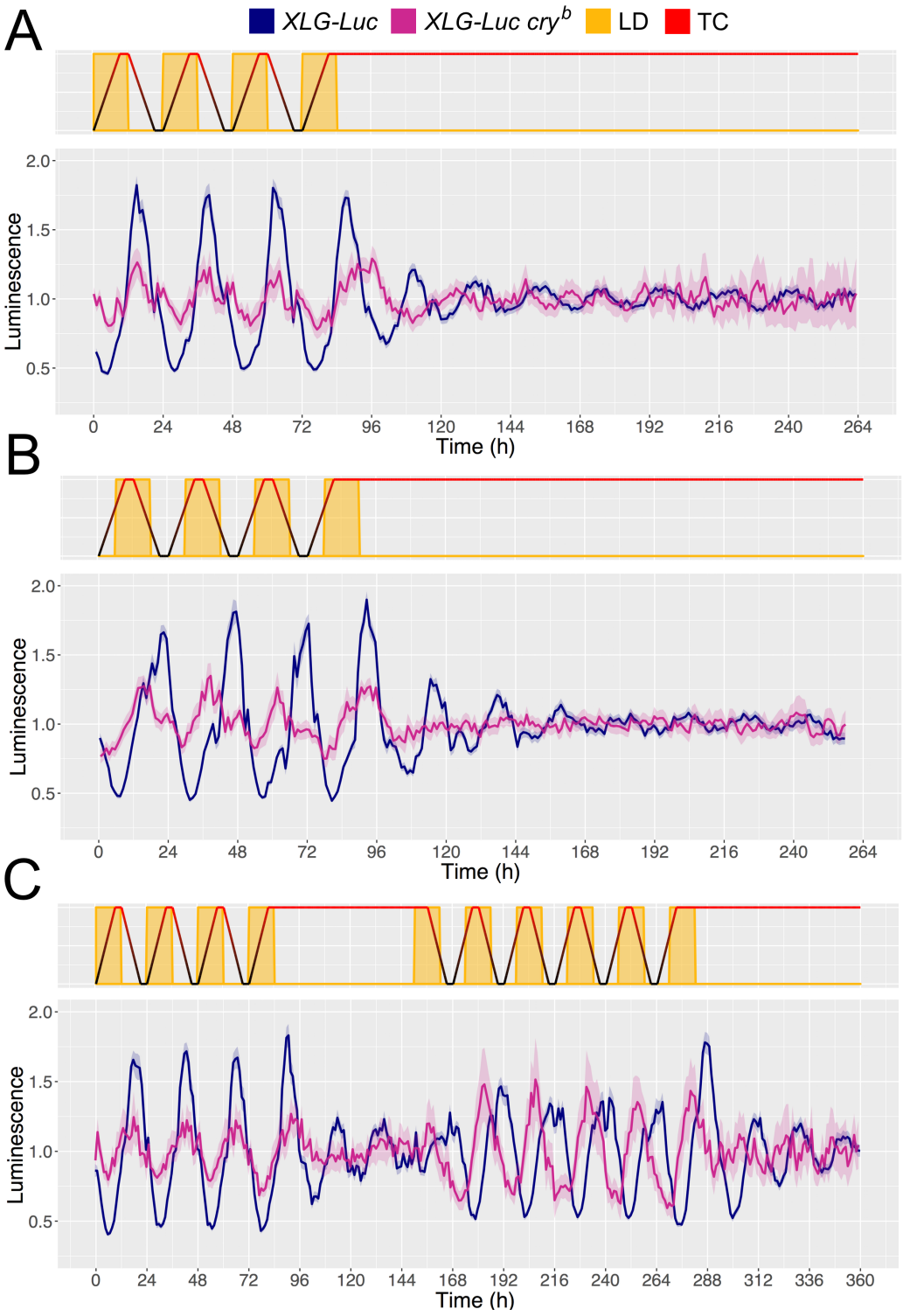
A
XLG-Luc
XLG-Luc cry^b
LD
TC
Luminescence
2.0
1.5
1.0
0.5
0
24
48
72
96
120
144
168
192
216
240
264
Time (h)
B
Luminescence
Time (h)
C
Luminescence
288
312
336
360
Time (h)

Figure 6.6: Bioluminescence recordings of peripheral clocks during sensory conflict. Dark blue and magenta lines show mean average of *XLG-luc* and *XLG-luc* cry^b flies respectively. Shaded regions show SEM. (A) Experimental regime in which environmental conditions followed 4 days of LD:TC in phase, and 7 days of free run in DD at 26°C. (n = 43-46 and n = 15-30 for wild type and *cry* mutants respectively). (B) Experimental regime in which environmental conditions followed 4 days of misaligned LD:TC, and 7 days of free run in DD at 26°C. (n = 47-48 and n = 28-35). (C) Experimental regime in which environmental conditions followed 4 days of LD:TC in phase, 2 days of free run in DD at 26°C, 6 days of misaligned LD:TC via 6 hr delay of LD, and finally 3 days of free run at DD and 26°C. (n = 37-44 and n = 7-14 for wild type and *cry* mutants respectively). All bioluminescence readings recorded at a resolution of 1 hr.

During 6 hr delay of LD relative to TC, peak bioluminescence in wild type flies was also delayed by approximately 6 hours (Fig. 6.6 B,C; Fig 6.7 left; Fig. A.1 A,B). However, a similar shift was not observed in the cry^b background, which instead showed no change in peak phase (Fig. 6.6 B,C; Fig 6.7 left; Fig. A.1 A,B). Together, these results indicate that peripheral PER rhythms in wild type flies entrain preferentially to light during conflicting LD:TC, whereas cry^b flies entrain preferentially to temperature. This directly contrasts results obtained in the central clock of the fly brain (see Chapter 5).

Strikingly, no significant effect of misaligned LD:TC was observed on the amplitude or rhythmicity in either wild type or *cry* mutant flies when pooling data across experiments (Fig. 6.7 middle, right). This again contrasts findings in the central clock neurons of wild type flies under similar environmental conflicts, in which the amplitude of PER oscillations was severely dampened during misaligned conditions (Harper et al., 2016). In fact, the amplitude of peripheral clock bioluminescence rhythms was consistently larger in wild type flies compared to cry^b mutants (Fig. 6.7 middle), contrasting observations made during conflict in the central clock neurons.

Owing to the presence of P behaviour in wild type flies, suggesting a similar breakdown of PER oscillations in the central clock neurons as described previously (Fig. 6.5 B,D; Harper et al. 2016), my results suggest that peripheral molecular clocks *do not* collapse during 6 hr misalignment of LD and TC, but rather syn-

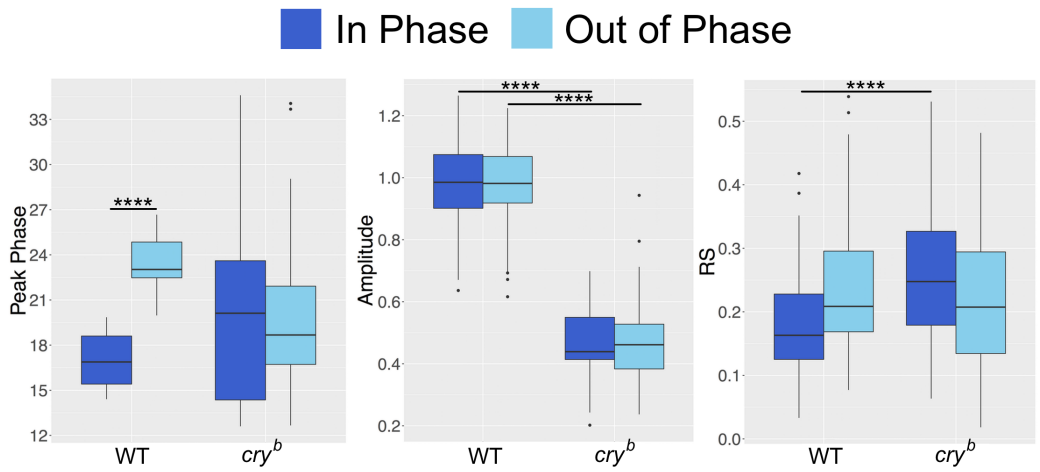

Figure 6.7: Quantification of peripheral bioluminescence rhythms during sensory conflict. Quantification of signal peak phase (left), amplitude (middle) and rhythmicity (right) during aligned and misaligned LD:TC. Analysis of the aligned condition used adult flies taken from the last full two days of aligned LD:TC (Fig. 6.6 A and C) (n = 80 and n = 22 for wild type and cry mutants respectively). Analysis of the misaligned condition used flies taken from the last full two days of misaligned LD:TC (Fig. 6.6 B and Fig. A.1 A) (n = 91 and n = 47 for wild type and cry mutants respectively). Box plots show 1st, 2nd and 3rd quartiles of the data, with the upper and lower whiskers extending to ±IQR from the 3rd and 1st quartiles respectively. Data points beyond the whiskers are plotted as outliers. Mann-Whitney U test used to compare between condition and genotype ($p < 0.00001$ shown by ****).

chronize to the light cue. In contrast, removal of CRY has similar consequences on both central and peripheral oscillators, rendering both preferentially sensitive to temperature cycles.

During free-running conditions, *cry*-negative flies became arrhythmic, consistent with the reported role of *cryptochrome* in the core clock machinery of peripheral clocks (Collins et al., 2006). In contrast, *cry*-positive flies continued to show bioluminescence rhythms after both aligned and misaligned conditions. These free running rhythms then gradually dampened with time (Fig. 6.6 A-C; Fig. A.1 A,B), consistent with previous studies in *XLG-luc* (Veleri et al., 2003). Median free running period of wild type flies after aligned and misaligned LD:TC was 21.5 hr and 22.0 hr, respectively, with no significant difference observed between these different entrainment conditions. This suggests no lasting effect of sensory conflict on peripheral clock rhythms. (An equivalent analysis for *cry* mutant flies was made

impossible by the complete lack of free running rhythmicity).

6.4 Discussion and Future Work

Disruptions of locomotor behaviour in wild type flies (leading to P behaviour) were shown in Chapter 5 to result from a 6 hr delay of LD relative to *square-wave* TC (Harper et al., 2016). Here, I provide evidence to suggest that evening locomotor behaviour is similarly altered during equivalent misalignments under more naturalistic temperature fluctuations. Indeed, the activity profile that results from a more naturalistic form of sensory conflict (using ramped TC) closely mimics the previously reported P behaviour, displaying a similar breakdown of evening anticipation, yet without the abrupt changes in activity observed previously at the end of thermo-phase and photo-phase (probably due to the lack of abrupt environmental transitions evoking masking behaviour).

While conflict between light and temperature caused behavioural disruptions in *XLG-luc* wild type flies, an associated disruption of molecular cycling in the peripheral clocks was not seen using a *luciferase* reporter assay: 6 hr misaligned LD:TC resulted in an equivalent phase shift of PER rhythms (i.e. peripheral clock rhythms remained synchronized to the light stimulus).

In further contrast to the central clock, the amplitude of PER oscillations was not significantly changed between aligned and misaligned conditions. Taken together, this strongly suggests that, during sensory conflict, peripheral clocks in fruit flies entrain preferentially, perhaps exclusively, to light. Peripheral clocks thus exhibit a separate, and distinct, response to sensory conflict compared to that observed in the central clock neurons. Future work would benefit from an investigation into how this response is affected by varying environmental phase relationships.

Another outstanding question is whether the peripheral clock has no effect on locomotor behaviour at all. Indeed, the finding that peripheral PER rhythms in control flies display no other change during sensory conflict than shifting their phase with the light cue, might indicate that the circadian anomalies observed during sensory conflict purely result from disruptions of the central clock. However, an alter-

native explanation could be that the P-like behaviour we observe emerges from a discrepancy between the peripheral and central clock networks. If peripheral clocks do contribute to locomotor behaviour, then activity under sensory conflict will necessarily be driven by two out-of-sync circadian networks. A potential route for future research would be to use the kinases *Doubletime* and *Shaggy* to generate period discrepancies between peripheral and central clocks, as was done previously to assess autonomy between neuronal subgroups within the central clock network (Yao and Shafer, 2014). Another option would be to simultaneously measure peripheral clock bioluminescence and activity in individual flies and investigate any subtle correlations therein (Guo et al., 2016; Khabirova et al., 2016).

The bioluminescence imaging data on page 97 show that *XLG-luc*, a previously used transgenic reporter line, is expressed primarily in the eyes and abdomen. Thus, the findings of this chapter predominantly relate to peripheral clocks located in these body parts. The peripheral circadian system, however, exhibits much heterogeneity, specifically with regard to the degree of independence from the central clock (Ito and Tomioka, 2016). It is not yet clear how my findings translate to other areas in the wider peripheral network. A similar rationale could be applied to the role of *cryptochrome* in peripheral clocks. My results show that light dominance in the periphery depends on *cry* expression; whether the *cry* dependence of this light dominance reflects a dependence within the peripheral clock itself or elsewhere in the circadian system, however, remains unclear. One approach to addressing this question would be to repeat the experiments presented here with isolated tissues samples cultured in Luciferin medium.

The question of why the peripheral clock network might respond differently to sensory conflict, when compared to the central clock, also remains unclear. The central clock is a highly interconnected network with strong coupling through the action of PDF (Lin et al., 2004). Thus, a dissociation between oscillatory components, resulting from sensory conflict, is a potential cause for the disruptions observed. Less is known about connectivity in the peripheral clock system. The resilience of these oscillators to sensory conflict may therefore hint at a more in-

dependent network architecture, with less coupling between subparts. This theory lends itself to modelling approaches; the type of network model discussed in Section 2.2.2 of Chapter 2, for example, might provide a useful framework to infer coupling strengths and guide further experimentation.

Equally, from a more Bayesian perspective, we might ask *why* the central clock does not appear to coordinate peripheral gut and eye clocks during conflict? Perhaps an uncertainty in the central clock oscillations, reflected in their reduced amplitude, is projected to the periphery in the form of a low precision signal, thus leading to more autonomous behaviour in these peripheral clocks. Such hypotheses could explain aspects of the heterogeneity observed throughout the wider clock network. I would here stress again the benefits to come from embracing these mathematical viewpoints, alongside more holistic experimental studies of circadian systems in multisensory environments.

6.5 Summary

In this chapter, I used a *per-luciferase* transgenic reporter to investigate the combination of circadian light and temperature cues in peripheral clocks of *Drosophila melanogaster*. To my knowledge, this is the first study of MSI in the peripheral clock network of the fruit fly.

My results show that the responses of peripheral clocks during conflicting sensory entrainment conditions differ markedly from those of the central clock, placing light as the dominant cue for these oscillators, and further highlighting the diversity within the wider circadian system. In this way, I hope to have addressed objectives 2 and 3 outlined in Chapter 3.

Chapter 7

A Peripheral Clock in the Fly Ear

The work presented in this chapter forms part of a wider collaborative investigation within the research group of Joerg Albert

7.1 Introduction

Having explored the concurrent processing of photic and thermal signals by both the central clock network (Chapter 5), and peripheral clocks in the gut and eyes (Chapter 6), let us now turn our attention to other known Zeitgebers in *Drosophila melanogaster*.

As discussed in Chapter 1, social interactions can synchronise daily locomotor rhythms in fruit flies (Fujii et al., 2007; Krupp et al., 2008; Levine et al., 2002c). This social entrainment has been shown to act via the transmission of odourant signals, revealing an olfactory input pathway to the *Drosophila* circadian clock. It is therefore especially interesting that olfactory sensation itself appears to be under circadian control: *per* oscillations are observed in chemosensory cells of the fly antenna (Tanoue et al., 2004), and daily rhythms have been reported in the electrophysiological responses of these cells to olfactory stimuli (Krishnan et al., 1999, 2008).

Additional to olfactory entrainment, locomotor activity in *Drosophila melanogaster* also synchronises to the rhythmic presentation of mechanical stimuli (see Chapter 1; Fig. 1.5; Simoni et al. 2014). Indeed, 12 hr:12 hr cycles of vibration followed by silence (VS) are sufficient to entrain daily activity rhythms in wild

type flies (Simoni et al., 2014). The VS stimulus also appears to entrain molecular rhythms in the fly brain, as shown by *per* oscillations in the central clock neurons (Simoni et al., 2014). Importantly, both behavioural and molecular rhythms were abolished in clock mutants, demonstrating their true circadian nature. However, entrainment was also abolished upon genetic ablation of chordotonal organs (ChOs), implicating these structures as key components of the mechanosensory clock input pathway.

ChOs are internal mechanoreceptors that mediate proprioception and the detection of air- and substrate-borne vibrations (Field and Matheson, 1998; Kernan, 2007). They are comprised of individual sensory units termed 'scolopidia' (Fig. 7.1 C), which are themselves multi-cellular stretch detectors, conserved across crustaceans and insects (Yack, 2004). The scolopidia contain dendritic projections of one or more afferent neurons, fused both apically (to the dendritic cap cells) and basally (via ligament cells) through microtubule interactions (Fig. 7.1 C). Scolopale cells form a capsule around the cilia of these sensory neurons, and produce ionic receptor lymph to promote electrical signalling. (For a full review of ChO structure and function, see Kernan 2007).

Three key transient receptor potential (TRP) channels have been identified as molecular components of *Drosophila* ChO mechanotransduction: the TRPN1 channel gene, *nompC*, and the mutually heterodimeric TRP vanilloid (TRPV) channel genes, *nanchung* (*nan*) and *inactive* (*iav*) (Goepfert et al., 2006; Kim et al., 2003; Walker et al., 2000; Zhang et al., 2013). While the precise roles played by these channels in mechanotransduction and amplification remain unclear (Albert and Goepfert, 2015), what *is* clear is that they constitute a vital contribution to fruit fly mechanosensation, highlighting them as viable genetic targets for investigations of mechanical input to circadian clocks.

In addition to the olfactory organs, the fly antenna also houses a key mechanosensory apparatus – the Johnston's Organ (JO) (Caldwell and Eberl, 2002; Todi et al., 2004). Located in the second antennal segment (a2), close to its chemosensory neighbour in the third antennal segment (a3), the JO comprises

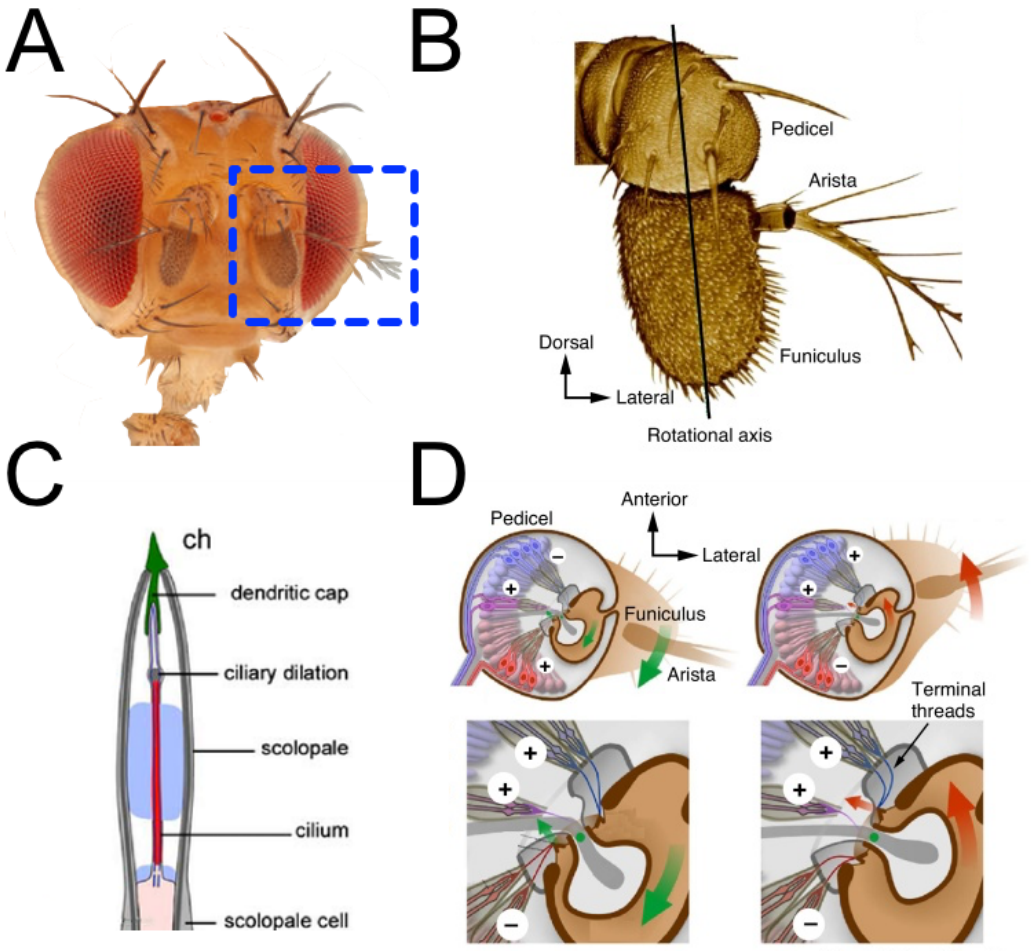

Figure 7.1: Structure and organisation of the *Drosophila* antennal mechanosensory apparatus. (A) Light microscopy image of the fly head. One of a pair of antennae is highlighted with a blue dashed box. (B) A diagram of a single antenna showing the three antennal segments, and the rotational axis between a2 and a3. (Adapted from Albert and Goepfert 2015). (C) A diagram showing the structure of a scolopidium containing a single neuron. (Adapted from Kernan 2007) (D) A diagram showing the internal structure of the second antennal segment containing the JO. Movements of a3 relative to a2 (shown by the green and red arrows) act to stretch the array of scolopidia subunits. (Adapted from Albert and Goepfert 2015).

~200 scolopidia, containing ~500 sensory neurons. The structure of the JO, and its position in the antenna is shown in Fig. 7.1 (A,B,D). The feathery 'arista' mechanically couples to the vibrations of sound waves, causing a3 to rotate around its longitudinal axis relative to a2. This rotation of the antenna along the intersection between the second and third antennal segments causes stretching of the scolopidia in the JO (Fig. 7.1 D), which in turn serves to activate or inhibit the sensory neurons therein. In this way, there exists a direct coupling between the mechanical forces in the surrounding environment, and stretch-sensitive neurons in the fly antenna (Caldwell and Eberl, 2002; Kamikouchi et al., 2009; Todi et al., 2004). For its role in detecting near-field sounds, the JO is often referred to as the fruit fly 'ear'.

Given the importance of locomotor activity as a key behavioural output of the *Drosophila* circadian system, it seems reasonable to suggest that a mechanosensory processing pathway is of particular interest for multisensory investigations of the fly clock. The proprioceptive nature of entrainment to mechanical signals in the external world (i.e. VS) also hints at a possible *internally*-maintained clock input pathway. A fly's own locomotor behaviour, which is bound to excite proprioceptive ChOs, might contribute to the final computation of internal time by the circadian network. In such a scenario, daily activity would represent both an input to, and output of, the fly clock, blurring conceptual boundaries, and providing the potential for complex temporal dynamics.

Taking this idea of ill-defined input-output distinctions further, complications are introduced through the potential for circadian modulations in the processing of ZG signals. Recall that olfactory signals in the environment act to entrain the *Drosophila* clock (Fujii et al., 2007; Krupp et al., 2008; Levine et al., 2002c); yet, olfaction itself appears to be under circadian control (Krishnan et al., 1999, 2008; Tanoue et al., 2004). This could provide a mechanism by which clock output actually gates its own input − such that the efficacies of timing cues are themselves adjusted in a time-dependent manner.

An exploration of this complex clock input-output from a mechanosensory perspective might begin by asking whether there exists a peripheral clock in the

JO of *Drosophila melanogaster*. In this chapter, I attempted to address this question in the hope of offering a foundation upon which future work might build. Drawing parallels with results in the fly olfactory system, I conducted an analysis of circadian gene transcription in tissue isolated from a2. Activity of the core clock genes (*per*, *tim*, and *Clk*) was quantified, as well as the mechanosensory genes, *nompC*, *nan*, and *iav*.

7.2 Materials and Methods

7.2.1 Fly Strains

Flies were reared under 12 hr:12 hr LD cycles on *Drosophila* medium (0.8% agar, 2.2% sugar-beet syrup, 8.0% malt extract, 1.8% yeast, 1.0% soy flour, 8.0% corn flour, and 0.3% hydroxybenzoic acid) at 25°C and approximately 60% humidity. Canton S flies were used as wild type.

7.2.2 Quantitative Gene Expression

Special thanks goes to Jason Somers (postdoctoral researcher in the Albert lab) and Ariana Huebner (visiting student) for their invaluable contribution to these experiments.

7.2.2.1 Dissections

Male wild type flies were entrained to 12 hr: 12 hr LD at 25°C. At 7 days old, flies were collected at 12 time points evenly spaced across the 24-hour day (corresponding to ZT 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22), and then immediately snap frozen in liquid nitrogen. Flies were next shaken to dissociate body parts, separating the heads and antennal segments. Both whole heads and a2 segments were collected and stored at -80°C. This dissection protocol had the benefit of cleanly separating a2 and a3, without the contamination risks involved in manual dissection using forceps (which was attempted in a pilot experiment).

7.2.2.2 Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

Three biological replicates were performed for both whole fly heads (40 per replicate) and a2 segments (60 per replicate). Tissue samples dissected as described previously were placed in lysis buffer containing β-mercaptoethanol and mechanically disrupted using a micropestle (*Eppendorf*). RNA was then purified from the tissue samples using the PureLink™ RNA Mini Kit (*Invitrogen*) with an added DNA digestion step using Purelink™ DNase to remove any contaminating DNA and further purify the RNA solution (*Invitrogen*). The RNA was then reverse-transcribed into cDNA using the High Capacity cDNA Reverse Transcription Kit (*Applied Biosystems*).

Due to the low amount of input RNA for the a2 sample preparations, pre-amplification of the cDNA was required prior to quantification. This was performed using the TaqMan™ PreAmp Master Mix Kit (*Applied Biosystems*), and included 12 amplification cycles. The final product was then diluted to 400 μl. No pre-amplification step was required for the whole head preparations due to sufficient levels of cDNA in these tissue samples.

cDNA quantification was performed using the StepOnePlus™ Real-Time PCR System (*Applied Biosystems*) and the TaqMan™ Gene Expression Master Mix (*Applied Biosystems*).

7.2.2.3 Gene Probes

A list of all gene probes and fluorescent reporters is shown in Table 7.1. The *rpl32* housekeeping gene was used as an endogenous control. To validate that the probes have similar amplification efficiencies over serial dilutions of the input cDNA (as is an assumption of the analysis method described below), RNA from whole fly heads (including antennae) was purified as described above. The relative quantification of the target probes compared to the endogenous *rpl32* control was shown to be consistent across varying dilutions (Fig. 7.2).

Gene probe	Ref. No.	Fluorescent reporter
rpl32	Dm02151827	VIC®
Clk	Dm01795382	FAM™
per	Dm01843681	FAM™
tim	Dm01814242	FAM™
nompC	Dm01808271	FAM™
nan	Dm01805137	FAM™
iav	Dm01833375	FAM™

Table 7.1: Gene probes

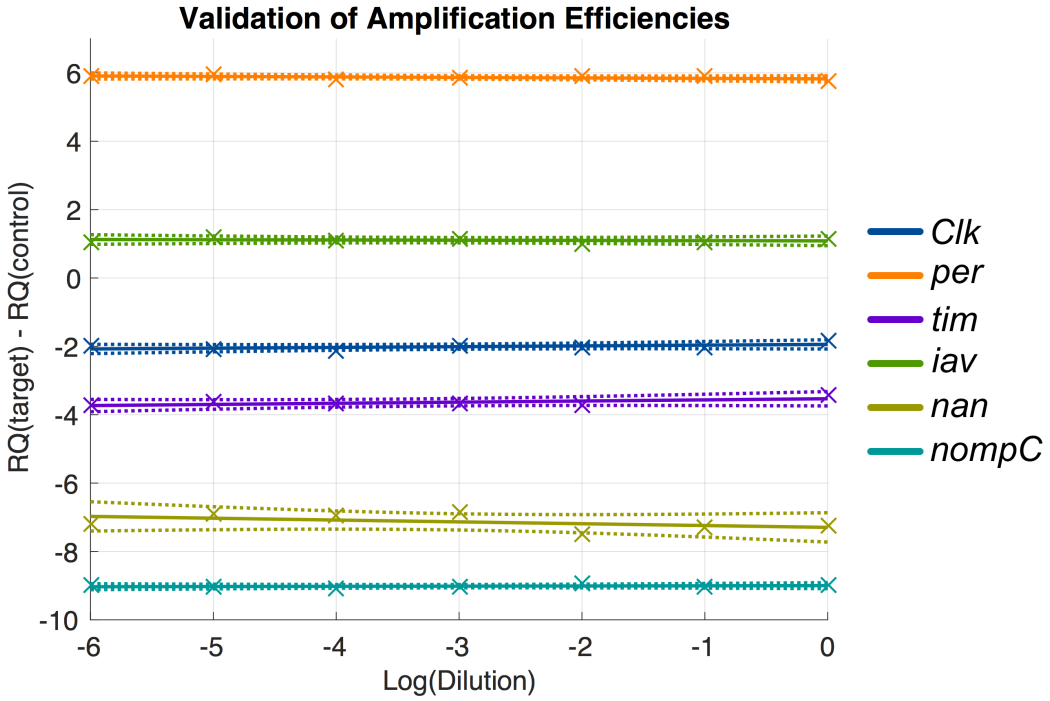

Figure 7.2: Validation of gene probes. RNA was extracted and purified from whole fly heads (n = 100). Relative quantification was measured using qPCR (see text) for different dilutions of the target sample. Mean RQ was calculated from three biological replicates for each RNA sample dilution. A linear model was fit using maximum likelihood, with parameters shown in Table. 7.2. Note the consistent RQ values across dilutions (linear gradient magnitude does no exceed 0.1; see Table. 7.2).

7.2.2.4 Analysis

Analysis of relative quantification was performed using the $2^{-\Delta\Delta C_T}$ method as described in Livak and Schmittgen (2001). Briefly, amplification of the target probe can be expressed as:

Gene probe	Linear fit gradient	Linear fit intercept
Clk	0.0022	-1.942
per	-0.014	5.825
tim	0.033	-3.533
nompC	0.005	-9.000
nan	-0.054	-7.295
iav	-0.007	1.086

Table 7.2: Linear fit parameters of log dilution vs. relative gene quantification

$$X_n = X_0(1+E_X)^n \tag{7.1}$$

where X_n is the number of target molecules at cycle n of the reaction, X_0 is the initial number of target molecules, E_X is the efficiency of target amplification, and n is the number of reaction cycles.

We can define C_T as the fractional number of reaction cycles required for the target to reach a threshold concentration. Similarly, ΔC_T is the difference in the cycle threshold between the target and the endogenous control ($C_{T_X} - C_{T_R}$). The initial target concentration, normalised against the initial endogenous control concentration, can then be expressed as:

$$X_{norm} = K(1+E_X)^{-\Delta C_T} \tag{7.2}$$

where K is some constant value.

Finally, dividing by a calibrator sample and assuming the efficiency, $E_X = 1$, K factors out and we get the final relative quantification:

$$RQ = 2^{-\Delta\Delta C_T} \tag{7.3}$$

where $\Delta\Delta C_T = (\Delta C_{T,q} - \Delta C_{T,cb})$, $\Delta C_{T,q}$ is the cycle threshold for the target sample,

and $\Delta C_{T,cb}$ is the cycle threshold for the calibrator sample.

7.3 Results

7.3.1 Transcriptional Rhythms of Clock Genes in Fly Heads

The transcriptional profile of the core clock genes in fly heads has been studied previously. Peak *per* and *tim* mRNA is typically measured in the early evening, around 6 hr prior to peak protein accumulation in the cytoplasm (Hardin, 1994; Sehgal et al., 1995). Conversely, *Clk* transcripts have been shown to oscillate in antiphase to *per* and *tim* during the early morning (Bae et al., 1998). While these peak expression times are broadly consistent across the literature, minor discrepancies are observed between different wild type genotypes, gender, and experimental procedures.

To control for these variations in the precise experimental design used, and to establish a clear picture of clock gene expression in the fly head, against which transcriptional activity in the JO would later be compared, qPCR was used to measure *Clk*, *per*, and *tim* mRNA in tissue isolated from whole fly heads, where the antennae had been removed (Fig. 7.3)

Consistent with previous findings, maximum *per* and *tim* mRNA was observed at ZT 16 (Fig. 7.3 orange and purple), with peak *Clk* transcripts measured 12 hr earlier at ZT 4 (Fig. 7.3 blue). This validates the experimental procedure used, and further provides a reference measurement of brain clock gene activity rhythms, which are expected to dominate in the whole fly head tissue samples.

7.3.2 Transcriptional Rhythms of Clock Genes in the Johnston's Organ

Having demonstrated sensible fluctuations of core clock gene transcripts in the fly head, qPCR was next used to measure *Clk*, *per*, and *tim* mRNA in tissue isolated from antennal segment a2, which houses the JO (see Introduction of this chapter; Fig. 7.1).

As was observed in the fruit fly head, peak *per* and *tim* expression was observed

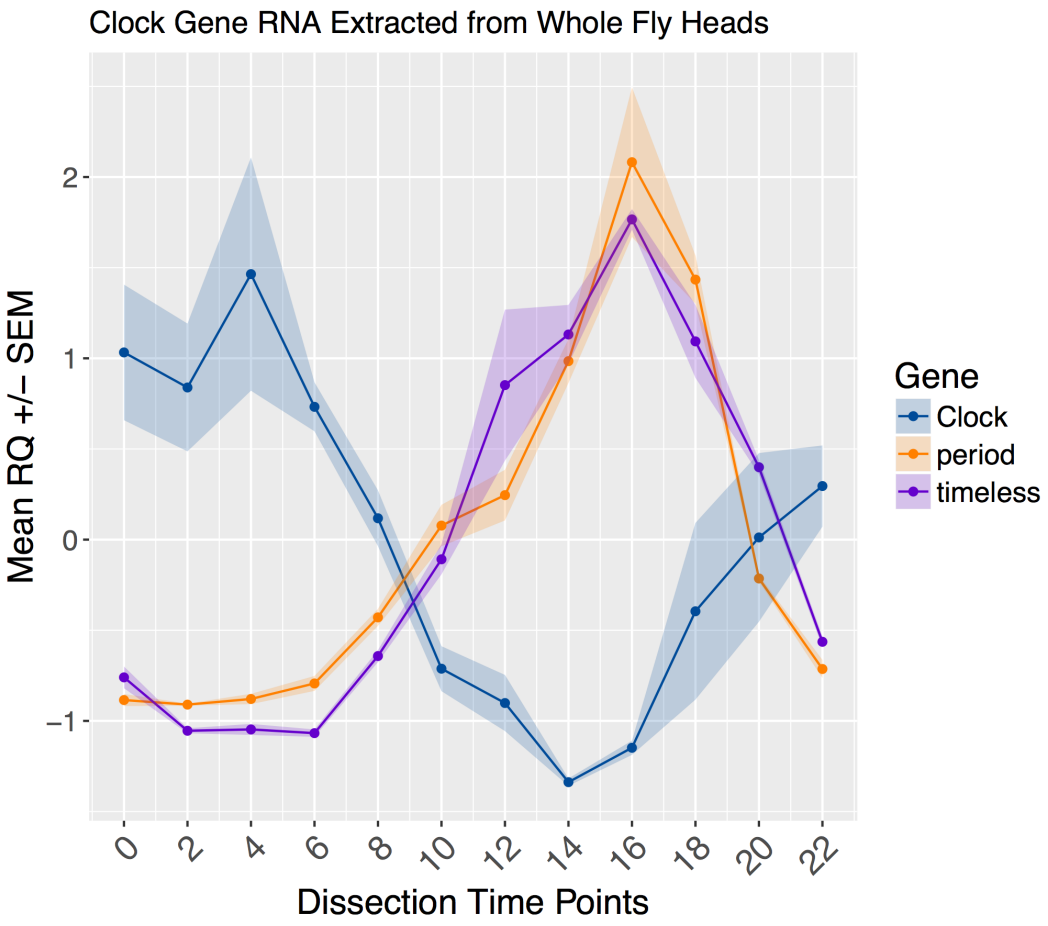

Figure 7.3: Cycling of the core clock gene mRNA in wild type fly heads. Flies were entrained to 12 hr: 12 hr LD cycles at 25°C and relative RNA quantified in tissue samples isolated at the time points shown (ZT). Each data point was calculated as the mean of three biological replicates (described in full on page 112), each comprising 40 fly heads with the antennae removed. Peak *Clk*, *per*, and *tim* expression measured at ZT 4, 16, and 16 respectively.

in the early night (Fig. 7.4 orange and purple). However, while maximum *per* mRNA was measured as before at ZT 16, peak *tim* was observed 2 hr later at ZT 18. Whether this discrepancy reflects a true difference between *tim* cycling in the brain versus the JO is unclear from the qPCR data alone. Indeed, special attention should perhaps be paid to these expression times in future studies.

Expression of the *Clk* gene also appeared to agree with that seen in the fly head, demonstrating broadly antiphasic rhythms to *per* and *tim* with peak mRNA measured during the morning (Fig. 7.4 blue). The data entry at ZT 6 deviates greatly from the pattern displayed in the remaining 11 time points, leading me to

suspect some form of anomalous reading. One explanation for this could be due to increased degradation of the RNA collected at ZT 6, the effects of which are seen particularly for the *Clk* quantification, as at this time, *Clk* was at near-maximal expression, whereas *per* and *tim* were already at their lowest. For this reason, it seems reasonable to omit the *Clk* mRNA measurement at ZT 6 from subsequent analysis and discussion.

Together, these data suggest the existence of a previously unreported peripheral clock in the *Drosophila* hearing organ.

7.3.3 Quantification of Mechanosensory TRP Channel Genes in the Johnston's Organ

Having identified peripheral rhythms of the core clock genes in the JO, I next investigated whether the three mechanosensory TRP channels, *nompC*, *nan*, and *iav*, displayed 24-hour cycles of transcription. Rhythms of these genes might indicate circadian modulation of mechanosensation, which would raise further questions on the interplay of clock output and input signals. qPCR was therefore used to measure mRNA of these three gene targets in tissue isolated from a2.

In contrast to the circadian profile of the core clock genes, neither *nompC*, *nan*, or *iav* displayed clear 24-hour cycling (Fig 7.5). In the absence of robust 24-hr oscillations, one-way analysis of variance was used to test for a significant effect of dissection time on the relative quantification of gene expression. Weakly significant fluctuations were observed for *iav* ($p = 0.026$), however this did not conform to the conventional sinusoidal rhythms observed for the core clock genes in the JO (Fig. 7.4). No significant effect was found for *nan* ($p = 0.182$) or *nompC* ($p = 0.366$). Thus, while circadian changes in *iav* expression cannot be ruled out, these data do not provide any clear evidence to suggest circadian control over the mechanosensory capabilities of the fruit fly. Future work is required to fully probe this hypothesis (see Discussion).

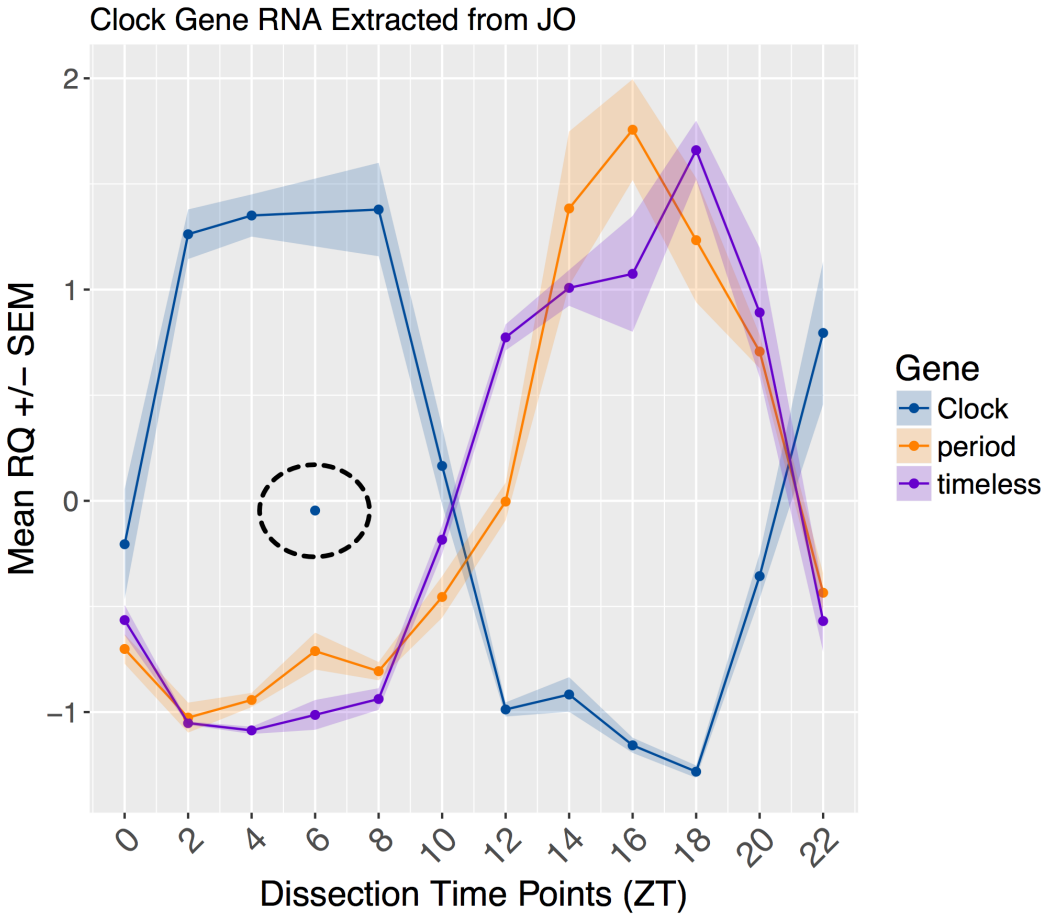

Figure 7.4: Cycling of the core clock gene mRNA in wild type a2 tissue. Flies were entrained to 12 hr: 12 hr LD cycles at 25°C and relative RNA quantified in tissue samples isolated at the time points shown (ZT). Each data point was calculated as the mean of three biological replicates (described in full on page 112), with each replicate comprising 60 a2 segments. Peak *per* and *tim* expression was measured at ZT 16 and 18 respectively. Peak *Clk* expression is unclear, owing to the presence of an anomalous reading at ZT 6, around where the trend in the data implicates peak expression to be. Note, broad transcriptional activity of *Clk* appears antiphasic to *per* and *tim*, agreeing largely with that seen in whole fly heads.

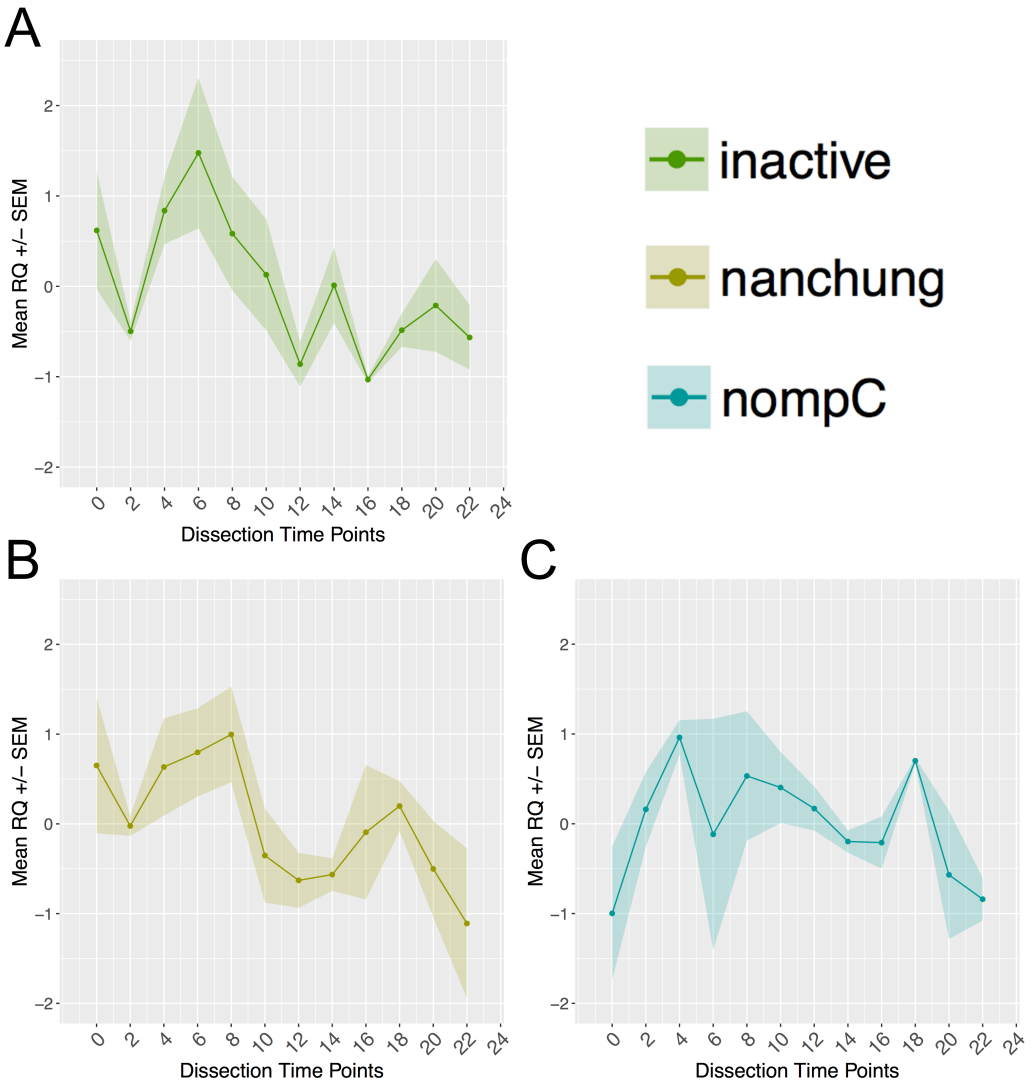

Figure 7.5: Quantification of mechanosensory gene mRNA in wild type a2 tissue. Flies were entrained to 12 hr: 12 hr LD cycles at 25°C and relative RNA quantified in tissue samples isolated at the time points shown (ZT). Each data point was calculated as the mean of three biological replicates (described in full on page 112), with each replicate comprising 60 a2 segments. Circadian rhythms of gene activity were not observed for any of the gene targets. This contrasts transcriptional oscillations in the core clock genes in this tissue (Fig. 7.4). One-way ANOVA on the data in A, B and C revealed $p = 0.026$ *, $p = 0.182$, and $p = 0.366$ respectively.

7.4 Discussion and Future Work

In Chapter 5 and 6, a multisensory treatment of the *Drosophila* circadian system revealed rich and complex forms of integration between photic and thermal timing cues. Here, I add to this sensory milieu by exploring the implication of an as yet lesser-studied mechanosensory input to the fly clock.

Inspired by previous work in the olfactory system, I sought to examine the potential for a peripheral circadian clock in the auditory system of *Drosophila melanogaster*. Using quantitative reverse transcription polymerase chain reaction (RT-qPCR), the relative quantification of core clock genes, and three key mechanosensory TRP channels, was assessed in tissue isolated from the second antennal segment (a2), which houses the JO − a crucial mechanosensory apparatus responsible for the detection of mechanical vibrations and sounds.

Strong oscillations of *Clk*, *per* and *tim* mRNA strongly suggest the existence of a peripheral clock in a2. An experimental challenge was presented by the close proximity of a2 to a3 (Fig. 7.1 A,B), which has been shown previously to contain a peripheral clock. To mitigate the risk of contamination between these two tissues, a specially-designed dissection protocol using liquid nitrogen to 'snap freeze' and dissociate body parts was used. This followed preliminary experiments involving a more conventional dissection of a2 with stainless steel forceps, which illustrated similar cycling of the core clock genes in a2 (data not shown).

While great care was taken to avoid cross-contamination of genetic material from a3, it should be noted that complete isolation of this tissue is made impossible by the inherent structure of the antenna. Afferent neurons from a3 must run through a2 *en route* to the fly brain, meaning that some tissue from this structure will inevitably contribute to the mRNA quantification. However, comparatively large amounts of tissue from a2 should mitigate these effects, which are likely to be absorbed into the background noise of the experimental protocol. Indeed, it would be very surprising that the robust clock gene rhythms observed in our data are attributable to contamination of the samples. Nevertheless, the data presented here would benefit from independent confirmation via alternative methods.

One such confirmation experiment would be to measure the levels of a reporter with expression restricted to a2. Here, *luciferase* assays of the type used in Chapter 6 present a powerful approach to measure daily fluctuations in protein targets. This approach is further desirable for its ability to quantify *in vivo* protein concentration, at high resolution, within a single animal (rather than pooling cytological staining data from multiple different flies). Coupled with advanced analysis tools (see Chapter 4), *luciferase* assays represent a gold standard in molecular chronobiology. However, the field is limited by the reporter constructs available for use. To my knowledge, there is currently no transgenic fly mutant expressing a *luciferase* fusion gene for either mechanosensory-specific targets, or clock gene targets restricted to the JO. Work is currently ongoing in our research group to remedy this state of affairs.

One possible route to mitigating the need for bespoke *luciferase* fusion genes might be to pair a *luciferase* reporter assay with the GAL4/UAS system in *Drosophila*. For example, crossing the existing mechanosensory drivers *nan-gal4* (also known as *F-gal4*; Kim et al. 2003), *iav-gal4* (Kwon et al., 2010), and *nompC-gal4* (Liu et al., 2007), with a *UAS-luciferase* line (Markstein et al., 2008), could permit bioluminescence recordings of *nan*, *iav*, and *nompC* respectively. In cases where the GAL4 driver is expressed in tissues other than the JO, provided it is not also expressed elsewhere in the antenna, one could feasibly examine the residual trace generated from subtracting data obtained in flies with their antennae removed, from those with intact antennae. This might isolate only the antennal contributions to the bioluminescence trace.

The qPCR data presented in this chapter further justifies a more extensive investigation of the a2 transcriptome through RNA sequencing (RNA-Seq). This would require dissection of tissue isolated from a2 samples at multiple time points across a 24 hr day (for which, a sensible protocol would be the one presented in this chapter). While this necessarily carries the same risks of contamination, RNA-Seq provides a more complete view of gene activity for a given moment in time. This would reduce the importance of identifying sensible targets *ad hoc*, facilitat-

ing a more agnostic investigation of gene cycling in the JO. Indeed, while robust circadian rhythms were not observed in the qPCR data for the TRP channels under investigation in this study, circadian modulation of mechanosensation could act via other gene targets.

Peripheral clocks are believed to be responsible for orchestrating local biological processes on a 24-hour schedule. The identification of clock gene rhythms in the JO presented in Fig. 7.4 thus has implications for circadian modulation of fruit fly mechanosensation. Indeed, this was a key motivation for the experiments conducted. A worthwhile next step, then, would be an in-depth functional analysis of the JO at different times of the day. Here, laser Doppler vibrometry (LDV) provides an exciting avenue to study the biomechanics of *Drosophila* hearing, providing exceptionally sensitive measurements of antennal displacements in response to either Brownian bombardment by air particles, or direct stimulation (Göpfert et al., 2005; Göpfert and Robert, 2003). Further mirroring studies in the *Drosophila* olfactory system (Krishnan et al., 1999), the LDV set-up can also be extended to perform electrophysiological recordings from the antennal nerve in response to mechanical stimuli administered at different times of the day. Similar experiments have been demonstrated previously (Albert et al., 2007).

The data presented in Fig. 7.4 provides evidence for a peripheral clock in the fly a2, however the extent to which this oscillator displays autonomy over the central clock remains unclear. Indeed, the peripheral clock system displays much heterogeneity, both in dominance relationships with the central clock, and in the role of CRY as a circadian photoreceptor and a core clock component (Ito and Tomioka, 2016). Moving forward, it would be interesting to define these features in the context of a peripheral clock in the fly ear. Functional studies of the kind described above might therefore be performed in a *per* 7.2:2 background, for which *per* oscillations (and therefore functional clocks) are restricted to the lateral neurons of the central clock (Frisch et al., 1994).

An in-depth investigation would also benefit from repeating the qPCR experiments presented in this chapter, using mutant flies where targeted ablation of spe-

cific clock network cell groups is carried out using the GAL4/UAS system. Potential candidates include *pdf-gal4* and *UAS-rpr* (Renn et al., 1999) to remove the PDF-expressing central pacemaker cells. Persistent *Clk*, *per*, and *tim* rhythms in the JO of these flies would indicate a strong level of autonomy in this peripheral clock. Finally, while I investigated the JO as a key proprioceptive organ, ChOs are of course present in other tissues around the fly body (Kernan, 2007). The experiments conducted in this chapter, and indeed many suggested in this discussion section, could thus be applied to other ChO-containing tissues. This might be necessary to further disentangle the role of the circadian clock in mechanosensation; and mechanosensation in clock entrainment.

7.5 Summary

Challenges of multisensory integration in the fruit fly circadian system extend beyond purely photic and thermal inputs. Mechanical stimuli also entrain the *Drosophila* clock (Simoni et al., 2014). This mechanical entrainment pathway raises interesting possibilities around how clock output might feedback on its own input. In this chapter, I used qPCR to measure transcriptional activity of core clock genes in the second antennal segment, housing the JO (a major mechanosensory organ). Robust circadian oscillations in *Clk*, *per*, and *tim* mRNA provides strong evidence for a previously unidentified peripheral clock in this tissue, which has implications for possible circadian modulation of mechanosensation, blurring the boundary between clock input and output.

Inspired by previous work in *Drosophila* olfaction, my results represent a firm foundation upon which future work investigating the mechanosensory properties of the fly clock might build. Indeed, functional circadian analysis of the JO is currently being performed by other members of the Albert lab. In this way, I hope to have addressed objective 4 outlined Chapter 3, and I look forward to further advances in this area.

Chapter 8

Statistical Modelling of the Circadian Clock

The work presented in this chapter is in preparation for publication.

8.1 Introductory Comment

In recent years, multiple industries have embraced machine learning (ML) for its powerful statistical capabilities. A recent report by the Royal Society highlights ML as a key disruptive force for the near future (The Royal Society, 2017). However, despite its undeniable potential, ML remains mostly absent from the field of chronobiology.

The work presented in this chapter constitutes a tentative revolt against the lack of ML and probabilistic modelling in circadian neuroscience. Necessarily, this demands mathematical discussions, which may be considered aversive to those with more experimental preferences. With this in mind, I have taken concerted efforts to maintain comprehension and relevance for the non-specialist, focussing on the key concepts, and relegating more lengthy derivations to the Appendix. Should this discussion at any point stray into overly-mathematical terrain for a given reader's comfort, I would encourage them to persevere. My goal, as ever, is to bridge the experimental and theoretical viewpoints. This carries a risk of offending both audiences; one that I take with enthusiasm.

8.2 Introduction

In Chapter 2, dynamical systems were introduced as the major modelling arm within the field of chronobiology. In accordance with this tradition, let us now consider one such deterministic model of the *Drosophila* clock, describing phase evolution in the presence of external entraining signals as a discrete dynamical system.

$$\theta_{t+1} = \theta_t + w + Z_t \phi(\theta_t) \tag{8.1}$$

Eq. 8.1 describes the phase of the clock in time, θ_{t+1}, in terms of its phase at the previous time-step, θ_t, and its intrinsic period, w (here equal to the size of the time step). A PRC term, ϕ, further describes the phase response of the clock to a Zeitgeber signal, Z_t, that takes binary values to denote the presence (1) or absence (0) of a Zeitgeber at time t. (To avoid confusion in this chapter, I will opt for the full spelling, 'Zeitgeber', in place of the 'ZG' shorthand used elsewhere). The PRC further depends on clock phase, in accordance with biological considerations of circadian systems (see Section 2.3 for a discussion of circadian PRCs).

In the absence of a Zeitgeber ($Z_t = 0$), Eq. 8.1 describes clock phase progression at a rate governed exclusively by the free-running period, w. When $Z_t = 1$, the rate of phase evolution is altered according to $\phi(\theta_t)$. Eq. 8.1 can thus be used to define a phase trace, from which we can go on to simulate locomotor activity by mapping values of clock phase to fly activity levels:

$$activity_t = f(\theta_t) \tag{8.2}$$

where $f(\theta_t)$ is some function that maps clock phase to expected activity levels. (I provide details for such a mapping function in Section 8.4.2).

A dynamical system model of this form could indeed display 24 hr free running period, be entrained by a Zeitgeber, and, given an appropriate choice for $f(\theta_t)$, generate realistic locomotor activity (Fig. 8.1; more details in Section 8.4).

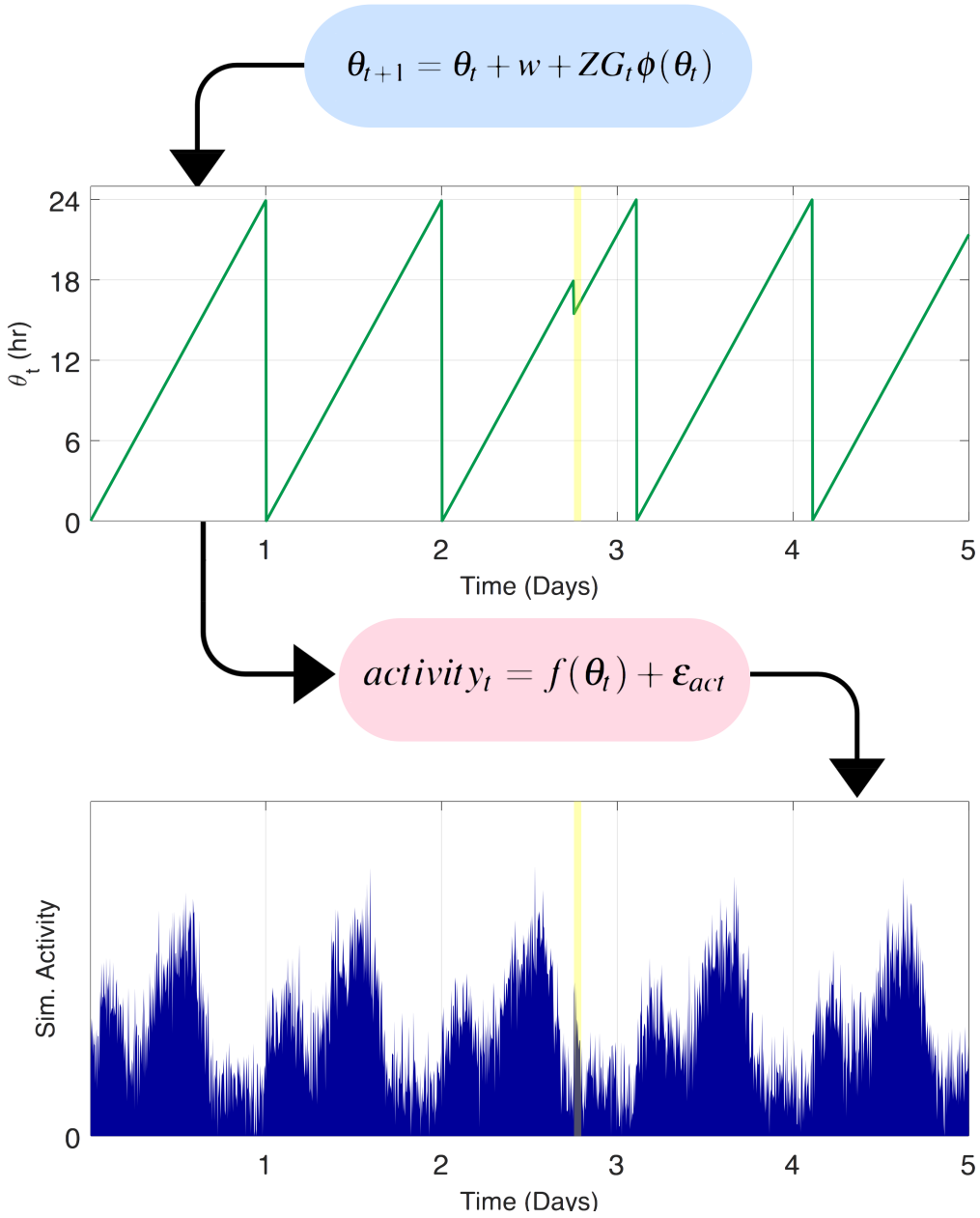

Figure 8.1: A simple model of the fly clock. A phase trace (green) is simulated from Eq. 8.1 (light blue) during a single Zeitgeber pulse (yellow). A simulated locomotor trace (dark blue) is then generated from this phase trace through some mapping function (pink). Further details can be found in Section 8.4

From the Marrian perspective outlined in Section 1.4.2, the model defined in Eq. 8.1 is an algorithmic-level rendition of the fly clock. True time of day is represented by some internal phase, which evolves according to an intrinsic frequency, and can be manipulated by a Zeitgeber signal. This model is deterministic in the sense that, given phase at time t, and the state of the entraining Zeitgeber signal, phase at $t+1$ can be calculated exactly. The future state of the clock is precisely determined through the known relationships in the model.

While the dynamical system model is an intuitive representation of the circadian clock, it is also perhaps a little restrictive. A more general formulation can be achieved through a purely probabilistic representation. Recall Marr's tri-level hypothesis of understanding, which teaches us that we should consider the computational problem that a neural system is solving, along with the implementation embodied in that system. This computational level of understanding provides a broader perspective on the underlying information processing issues at stake.

For the case of a circadian system, there is a true, objective time of day that the animal must infer to behave appropriately. There are cues to the time of day − the Zeitgebers − whose form is tied to this objective time; but also the prior information that day length tends to be constant.

Paralleling the animal's task of inferring objective time of day, and behaving accordingly, the experimenter's job is to infer the *subjective* time of day that the animal currently believes it to be. This is known as a *latent* or hidden variable, since it is not directly observable (see Section 1.4.2). To infer this latent variable, the experimenter can build a model of the animal's clockwork; can use their own observations of (or interventions on) the Zeitgebers, which they know the animal experiences; and can also observe the activity of the animal, which is tied to its subjective experience of time.

In this way, there exists a close coupling between the experimenter's and the animal's conceptions of time-keeping. I will therefore discuss these together, albeit focussing on the experimenter's view. For both the animal and the experimenter, the problem structure can be usefully seen as a hidden Markov model (HMM) − a class

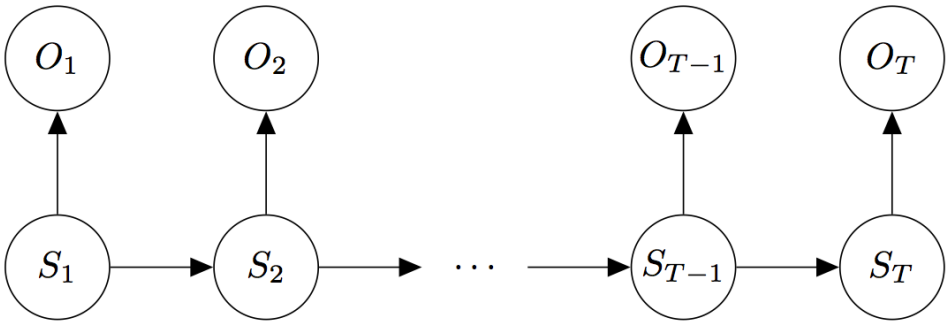

Figure 8.2: Graphical representation of a hidden Markov process. State transitions at each time step generate a sequence of hidden states, **S**, that in turn generates a sequence of observable values, **O**.

of generative latent variable model that represents discrete probability distributions over a sequence of observations.

The HMM is a well-established and widely-used latent variable model in ML, computational neuroscience, and across scientific disciplines. An extension of the work pioneered by Russian mathematician, Andrey Markov, HMMs are a ubiquitous tool for modelling time series data. Under this framework, an observation at time t, $O_t \in \mathbf{V}$, is stochastically generated according to an 'emission probability' by some underlying process, whose state, $S_t \in \mathbf{Q}$, is hidden from us. For us, the set of all states $\mathbf{Q}$ is the set of all possible times of day (which we ultimately discretize to one per half an hour). At each discrete time step, the system makes a transition from one hidden state to another (possibly a transition back to itself) according to some pre-defined 'transition probability'. We can represent this chain of events graphically, as shown in Fig. 8.2.

According to this abstraction of the clock, internal representations of time evolve *stochastically*, with these statistical properties defined by the transition probabilities of the HMM. Contrast this with the deterministic process described in Eq. 8.1. Likewise, mapping from the internal representation of time (S_t) to locomotor output (O_t) is a stochastic process captured by the emission probabilities in the model (akin to the role of Eq. 8.2 in the dynamical system formulation).

The graphical model depicted in Fig. 8.2 can free-run, but lacks a means for a Zeitgeber to influence phase. It would be interesting to adopt the perspective of the animal, in which the Zeitgebers are statistically considered to be *outputs* caused by the objective time of day (with Bayes rule allowing the former to influence the latter through the process of inference). However, here we adopt the simpler data-analytical scheme known as an input-output hidden Markov model (IOHMM; Bengio and Frasconi 1995). In this, the entraining effect of the Zeitgeber is exerted through alteration of the transition probabilities between the hidden clock phase states. In a similar way, we might also model behavioural masking (a phenomenon outlined in Chapter 1) as a direct effect of the Zeitgeber on locomotor activity, occurring through changes in the emission probability structure.

At the time of writing, I am aware of one other notable application of HMMs to chronbiology, where the standard HMM framework was used to infer circadian phase from *Rev-Erb*α-YFP fluorescence data in mammalian fibroblast cells (Bieler et al., 2014). The results of this study are promising, and provide firm justification for the use of HMMs in a circadian context. However, Bieler et al. (2014) stopped short of a full probabilistic model, instead using a parametrised harmonic function to describe the waveform of the oscillatory signal (see page 40 for issues associated with such an approach). Furthermore, use of a standard HMM may not have provided a complete descriptive model of the clock.

In this chapter, a novel IOHMM of the *Drosophila* circadian clock was designed and constructed. Through biological considerations of circadian systems, the transition probability structure of the model was constrained to a newly-defined class of 'circadian architectures'. The resulting probabilistic model can be considered a mathematical generalisation of previous dynamical system models, and is a closer representation of the computational task that must be solved by the fly clock.

The IOHMM was capable of learning circadian response dynamics from simulated and experimentally-derived data, and was able to perform statistical inference of hidden clock phase on locomotor time series, demonstrating its immediate potential as a tool for advanced phase analysis within the circadian field.

8.3 An Input-Output Hidden Markov Model of the Circadian Clock

Two assumptions are made about the trajectory of latent states, **S**, that allow us to reason tractably about the system described in Fig. 8.2:

1. There is a **limited horizon**, such that the probability of being in a given state at time t is dependent only on the state at time $t-1$. Formally:

$$P(S_t|S_{t-1}, S_{t-2}, ..., S_1) = P(S_t|S_{t-1}) \tag{8.3}$$

   This limited horizon assumption is often referred to as the 'Markov property', and the conditional independence relations are illustrated in Fig. 8.2.

2. The process is **stationary** in that the conditional distribution over the next state, given the current state, does not change over time.

Taken together, these assumptions allow us to write the joint probability distribution of the time series shown in Fig. 8.2 in the following factored form:

$$P(S_{1:T}, O_{1:T}) = P(S_1)P(O_1|S_1)\prod_{t=2}^{T} P(S_t|S_{t-1})P(O_t|S_t) \tag{8.4}$$

A full tutorial on HMMs goes beyond the scope of this thesis, however there are many excellent resources on the subject, of which my personal favourite is Rabiner and Juang (1986) for its gentle pace and intuitive explanations.

8.3.1 IOHMM Definition

Having outlined the standard HMM framework, let us now formally define the model in the context of the circadian clock.

T: length of the observation sequence

N: number of hidden clock phase states in the model (here 48 hidden states were used to represent 30 min resolution along a 24 hr phase cycle)

M: number of observable activity states in the model (here chosen to be 30 distinct levels of locomotor activity)

$\mathbf{V} = \{v_1, v_2, ...v_M\}$: the M potentially observable locomotor activity states

$\mathbf{O} = \{O_1, O_2, ...O_T\}$: the observed locomotor activity time series

$\mathbf{Q} = \{q_1, q_2, ...q_N\}$: the 48 hidden states representing clock phase (with $q_1 = 00:00, q_2 = 00:30, \ldots$)

$\mathbf{S} = \{S_1, S_2, ...S_T\}$: the hidden clock phase sequence

$A(Z) = \{a_{ij}(Z)\}$, where $a_{ij}(Z) = P(S_{t+1} = q_j | S_t = q_i; Z)$: a matrix of the hidden state transition probabilities. In the IOHMM, these can depend on the state of the Zeitgeber, Z, to account for the phase response curve.

$B(\Delta Z) = \{b_{ik}(\Delta Z)\}$, where $b_{ik}(\Delta Z) = P(O_t = v_k | S_t = q_i; \Delta Z)$: a matrix of the observable state emission probabilities. In the IOHMM, these can depend on transitions in the state of the Zeitgeber, ΔZ, to account for masking.

$\pi = \{\pi_i\}$, where $\pi_i = P(S_1 = q_i)$: the initial hidden phase state distribution

λ: the combined model parameters, $A(Z)$, $B(Z)$, and π

8.3.1.1 The Three Classic Problems of HMMs

Three main problems must be addressed in order to effectively employ an HMM:

1. The **evaluation problem**. Given an HMM, its parameters λ, and an observation sequence $\mathbf{O}$, what is the probability that the observation sequence was generated by the HMM? Formally, how do we calculate $P(\mathbf{O}; \lambda)$?

   For the animal, this reports how well its model of the environment fits the observations (the Zeitgebers). For the experimenter, it reports how well our model of the animal's clock fits the observations (the locomotor data).

2. The **learning problem**. Given the structure of an HMM, and one or more sequences of observations $\mathbf{O}$, how should we adjust the model parameters λ to maximise $P(\mathbf{O}; \lambda)$? This is trying to optimise the model fit to the data.

3. The **decoding problem**. Given an HMM, its parameters λ, and an observation sequence **O**, what is the most likely sequence of hidden states that produced the observations?

   This cleanly describes the problem for the experimenter – given all the observations of locomotor activity in the present and the past, infer what is the animal's current subjective time of day (as represented by the phase of its circadian clock). There exists a similar problem from the animal's perspective, except that the animal faces an additional issue of causality, only observing previous Zeitgeber values to work out the current objective time.

A full description of these problems goes beyond the scope of this thesis. Importantly, solutions are provided by the forward-backward, Baum-Welch, and Viterbi algorithms respectively (see Rabiner and Juang 1986).

8.3.2 A Novel Circadian Transition Probability Structure

The reader will recall from Chapter 1 that, in order for a system to be classified as circadian, it must display ~24 hr free running rhythms, while also possessing the capacity for entrainment to an external Zeitgeber signal. In an HMM, these circadian traits manifest themselves in the hidden clock phase space, the dynamics of which are defined by the transition probabilities in matrix $A(Z)$.

8.3.2.1 Existing Architectures

Hidden state dynamics are determined exclusively by the transition probabilities of an HMM. This statistical structure can take different forms. An ergodic (fully-connected) architecture is one where every state can be reached from all other states (Fig. 8.3 left). Such models may be able to fit circadian data, however they might also ignore biological plausibility. Indeed, free fitting of these transition probabilities could, for example, permit large and erratic transitions in the phase space, so long as this provides an explanation of the data. However, we know this explanation is likely to be incorrect, as we do not observe similar properties in circadian systems.

Alternatively, a 'left-right' architecture specifies sequential transitions in one

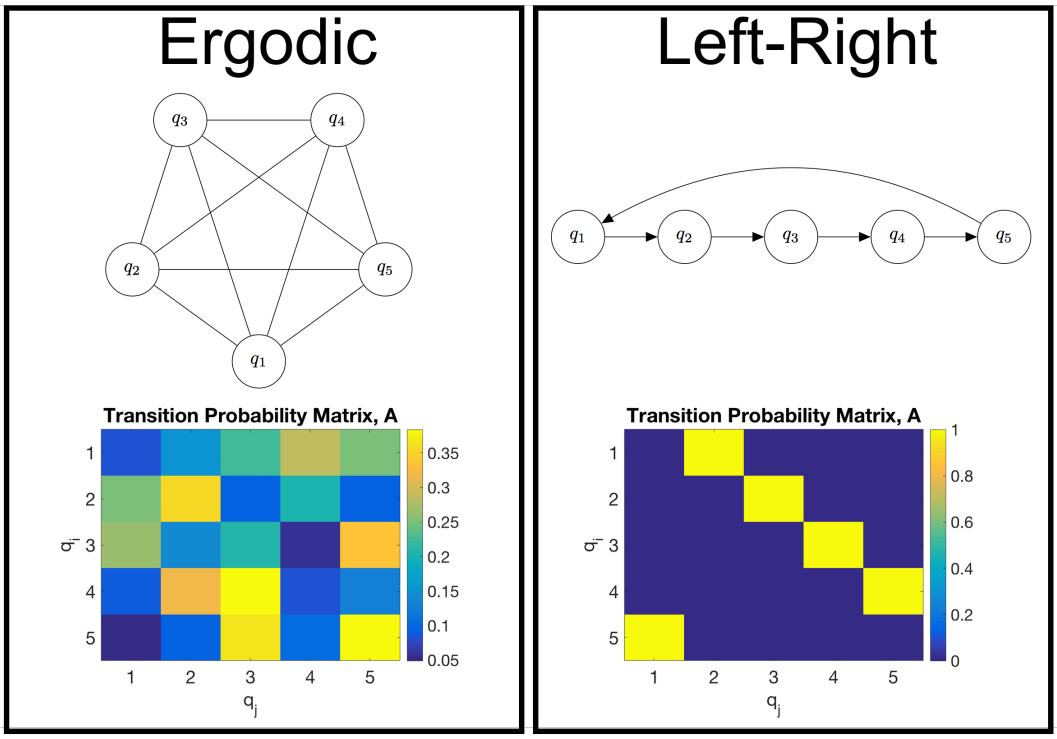

Figure 8.3: Common HMM architectures. Graphical representations of an ergodic HMM (top left) and a left-right HMM (top right), accompanied by representative transition probability matrices (bottom left and right respectively). In the ergodic architecture (left), transitions from one state to all others is a possibility. Learning in such a model can produce a 'chequerboard' transition probability matrix such that state sequences display erratic properties. In the left-right architecture (right), hidden state transitions are only permitted sequentially and unidirectionally along the chain. Due to probabilities of either 1 or 0, learning of the transition probabilities does not take place.

direction along a chain of hidden states (Fig. 8.3 right). We might consider this a more reasonable model of clock phase progression, which generally tends to move in a step-wise fashion. With the inclusion of a chain-closing transition from the last state back to the first (as shown in Fig. 8.3 right), a strict cyclic structure is imposed, which could well capture 24-hour periodicity. However, unlike circadian systems, such strict models cannot be entrained. Without more freedom to adjust phase evolution in time, a left-right HMM is doomed to eternal free run.

8.3.2.2 Proposed Circadian Architecture

To construct an HMM that honoured the aforementioned clock principles, I developed a novel 'circadian' architecture. One can think of this as a blend of the

structures mentioned previously: either as an ergodic model that tends to make transitions from left to right, or a left-right model that remains capable of larger transitional jumps when necessary. More formally, this circadian architecture obeys the following rules:

1. The most likely transition is q_i to q_{i+1}, with a chain-closing transition from q_N to q_1
2. It is possible to make a transition from q_i to q_j, for all $j \in 1,2,\ldots N$
3. The probability of a transition from q_i to q_j becomes very small as $\frac{N}{2} - ||i-j| - \frac{N}{2}|$ grows

Imposing these rules required adjustments to the standard HMM framework. Rather than treating each entry in matrix $A(Z)$ independently, the transition probabilities between hidden states were instead generated by a kernel function, k_A:

$$k_A(i,j) = a_{ij} = (1-C)e^{-\tau^2_{i_{del}} g(i,j)} + Ce^{-\tau^2_{i_{adv}} g(i,j)} \tag{8.5}$$

where:

$g(i,j) = \frac{N}{2} - ||i-j| - \frac{N}{2}|$

$C = 1$ when $j > i$, and $C = 0$ otherwise

$\tau_{i_{del}}$ is the decay rate for the delay transition probability

$\tau_{i_{adv}}$ is the decay rate for the advance transition probability

This kernel function deserves closer examination. Conceptually, it plays a similar role to the squared-exponential kernel used elsewhere in the field of Machine Learning (particularly Gaussian Process modelling; see Section 8.4.2). It imposes a structure on the relationship between variables – here, the transition probabilities between hidden states. However, whereas a squared-exponential kernel employs a distance metric of $(i-j)^2$ between matrix indices, $k_A(i,j)$ uses a circulant distance metric $g(i,j)$ to generate a transition probability matrix where $1 \rightarrow 12$ hr hidden state

transitions are considered the largest. This captures the periodic structure of clock phase.

Furthermore, this transition probability kernel comprises two terms: one that applies when advancing clock phase (with an associated rate parameter, $\tau_{i_{adv}}$); and one that applies when delaying clock phase (with rate parameter $\tau_{i_{del}}$). Which term contributes to evaluation of the transition probability is governed by the indicator variable, C, which is 1 when transitioning along the sequence and 0 when transitioning backwards. Note, for ease of reference in the text, 'τ' can be used to denote all τ parameters (both $\tau_{i_{del}}$ and $\tau_{i_{adv}}$).

It is clear from Eq. 8.5 that as the distance $g(i,j)$ between two phase states grows, the probability of a transition from q_i to q_j decays exponentially at a rate governed separately for each state, as well as separately for both the advance and delay directions. When all τ values are large, the HMM tends towards a purely left-right architecture (Fig. 8.4 B right); smaller τ values permit larger jumps in the phase space (Fig. 8.4 B left). In this way, the transition probability kernel function, $k_A(i,j)$, restricts HMM architectures to those that display circadian dynamics, maintaining biological relevancy in my model.

8.3.3 Modelling the Effect of a Zeitgeber on Clock Phase

In a circadian context, input to the HMM can be thought of as the value of a Zeitgeber signal sampled at discrete time points. This input sequence can then act to transform the transition probabilities between hidden clock phase states (Fig. 8.5, blue arrows; Eq. 8.6). This constitutes an IOHMM.

Implementation of the IOHMM architecture was achieved by generating the decay rate parameters, τ, of the transition probability kernel described in Eq. 8.5, in a Zeitgeber-dependent manner:

$$\begin{aligned} \tau_{i_{del}} &= \mu_{i_{del}} + Z_t \kappa_{i_{del}} \\ \tau_{i_{adv}} &= \mu_{i_{adv}} + Z_t \kappa_{i_{adv}} \end{aligned} \tag{8.6}$$

where Z_t is a binary variable that denotes the presence (1), or absence (0), of a

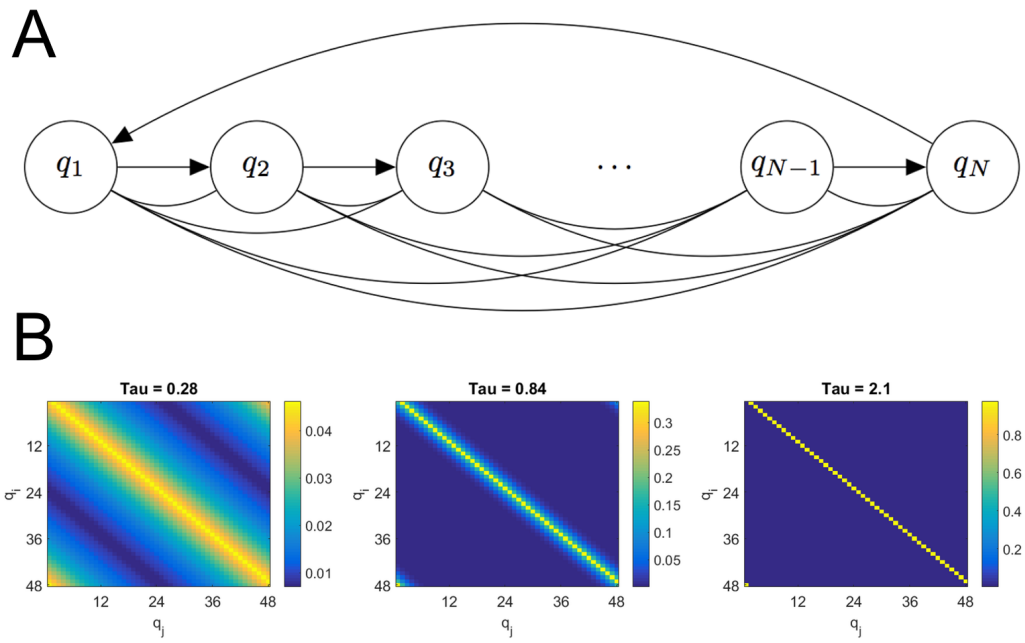

Figure 8.4: A circadian phase transition probability structure. (A) Graphical representation of my proposed 'circadian' transition probability structure. Note the ergodic (fully connected) structure making all transitions possible (shown by edges beneath the nodes), combined with a tendency to transition from left to right along the phase state chain (shown by the arrow-headed edges between nodes). The phase state chain is closed by a transition back from q_N to q_1. (B) Transition probability matrices generated by the kernel function described in Eq. 8.5, using different τ values: $\tau = 0.28$ (left), $\tau = 0.84$ (middle), and $\tau = 2.1$ (right). $\tau_{i_{del}}$ and $\tau_{i_{adv}}$ were set equal for all i in these depictions; this does not need to be the case in practice. Colour ranges from dark blue to yellow showing small to large transition probabilities (see scale bars). Note $k_A(i, j)$ always guarantees the most likely transition is to the next state in the chain, and that when τ is small, the system is more likely to make larger phase state jumps than when τ is large. In the latter case, my circadian transition probability structure approximates a left-right architecture.

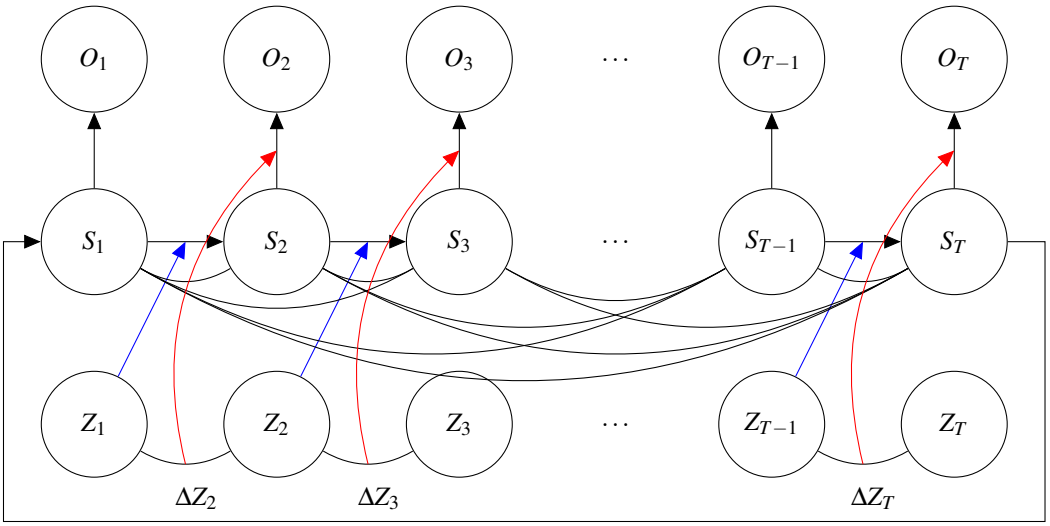

Figure 8.5: A circadian IOHMM architecture where a Zeitgeber influences both clock phase dynamics and activity levels. Graphical representation of an IOHMM implementation, where Zeitgeber input at each time step acts to alter the transition probabilities (blue arrows) between hidden states. For ease of comprehension, blue arrows are shown only for sequential transitions, however note that Zeitgeber input alters the entire probability structure of the model (including larger phase jumps). When $Z_t = 1$, the κ parameters contribute to the transition probability calculation, and when $Z_t = 0$, clock phase dynamics are governed exclusively by the μ parameters (Eq. 8.6). Additionally, a change in the Zeitgeber input between two time steps (ΔZ_t) acts to alter the emission probabilities (red arrows; Eq. 8.7). In this way, the model can capture masking behaviour, while preserving the statistical structure of the circadian waveform.

Zeitgeber signal at time t. Accordingly, the μ parameters govern the transition probability structure during free run, and the κ parameters adjust these statistical properties during presentation of a Zeitgeber.

Note that this introduces a time-dependency to the HMM. Indeed, as the conditional distribution over the next hidden state in the sequence *can* change in time, one could argue that this violates the assumption of stationarity outlined on page 131. The IOHMM remains structurally stationary, however, and the Markov property still applies.

8.3.4 Modelling Behavioural Masking

While much attention has been paid to circadian mechanisms behind Zeitgeber modulation of locomotor rhythms, it is important to recognise the existence of other non-circadian effects on fruit fly activity. Defined in Table. 1.2, and discussed

throughout this thesis, behavioural 'masking' can be observed in *Drosophila* activity patterns, particularly the pronounced startle during abrupt environmental transitions (i.e. Zeitgeber onset and offset). A good statistical model of the clock system must therefore seek to capture this startle behaviour.

To model these effects in my IOHMM, a second ΔZ input was derived from the existing Z input. This discrete time sequence took value 1 (when a change in the Zeitgeber sequence occurred; $Z_t \neq Z_{t+1}$), and 0 (when the Zeitgeber sequence remained constant; $Z_t = Z_{t+1}$). To avoid drastic alterations to the emission probability structure of the IOHMM, which is primarily fit on the relative abundance of data where no change in Zeitgeber occurs, $\Delta Z = 1$ acted to shift the mode of the emission probability distributions for each row of matrix B, by a factor of η (floored to the nearest integer).

$$b_{ik}(\Delta Z_t) = (1 - \Delta Z_t) b_{ik} + \Delta Z_t b_{i(k - \lfloor \eta b_i(\hat{k}) \rfloor)} \tag{8.7}$$

where $b_{ik}(\Delta Z_t)$ is the mode of the emission probability distribution for hidden state q_i.

It is clear from the above equation that when the Zeitgeber stimulus is constant ($\Delta Z_t = 0$), the emission probabilities in the IOHMM are unchanged; but when there is a change in the Zeitgeber ($\Delta Z_t = 1$), probability mass is shifted over higher levels of activity by factor η (for $\eta > 0$). Examples of these transformations are illustrated in Fig. 8.6. Note that η scales the mode of the distribution, leading to larger shifts in distributions with most of their mass over higher activity levels. This was designed to capture the notion of arousal, where startle responses tend to be exaggerated at times when the flies are more active (Van Swinderen and Andretic, 2003).

By temporarily adjusting the emission probabilities upon changes in the Zeitgeber signal, in a manner that preserves their shape, I aim to retain the rich statistical structure of the free running activity waveform, while also modelling the sharp, transient locomotor changes in response to environmental transitions.

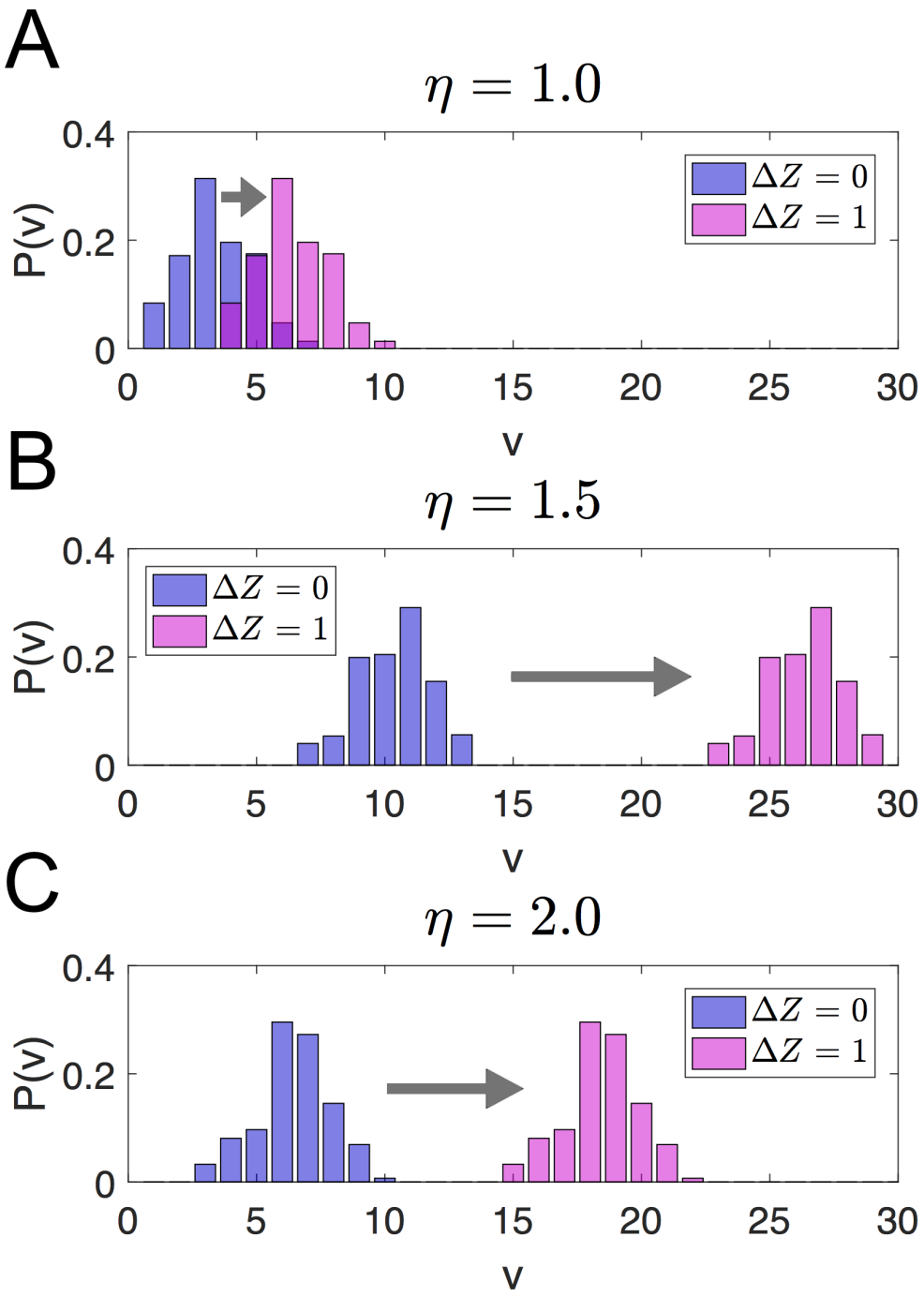

Figure 8.6: A ΔZ-dependent transformation of discrete probability distributions to capture startle behaviour. The transformation defined in Eq. 8.7 on three discrete probability distributions over activity states, v, for $\eta = 1.0$ (A), $\eta = 1.5$ (B), and $\eta = 2.0$ (C). The discrete probability distributions depicted are equivalent to individual rows of the IOHMM emission probability matrix, B. When $\Delta Z_t = 1$ (pink), the probability distributions are right-shifted compared to when $\Delta Z_t = 0$ (blue). Note that shape of the distributions is preserved during transformation. In middle and bottom, an increase in η does not increase the degree of shift due to the mode of the distribution in bottom being over a lower activity state. In this way, Eq. 8.7 captures the notion of arousal – startle responses are exaggerated in flies that are more active.

8.3.5 Learning the Parameters of the IOHMM

8.3.5.1 Expectation Maximisation

Given a set of observations, fitting the parameters, λ, of an HMM is referred to as the learning problem (see page 132). The standard solution to this problem implements the Baum-Welch algorithm, which was first described by Leonard Baum in the late 1960s (Baum et al., 1970), and is a particular instantiation of expectation-maximisation (EM) used widely in the field of Machine Learning.

EM is a general method for finding the maximum-likelihood estimate of the parameters of an underlying distribution, from data that is incomplete or has missing values. Recall the standard definition of the likelihood function:

$$\mathcal{L}(\lambda;\mathbf{D}) = P(\mathbf{D};\lambda) \tag{8.8}$$

where $\mathcal{L}$ is the likelihood function, $\mathbf{D}$ are the data, and λ is the set of model parameters.

We see that the likelihood is simply the probability of the data under the model parameters. Maximum likelihood seeks to find the parameters λ^* that maximise this likelihood:

$$\lambda^* = \underset{\lambda}{\operatorname{argmax}}\, \mathcal{L}(\lambda|\mathbf{D}) \tag{8.9}$$

In the case of an HMM, the data are the observations $\mathbf{O}$; the hidden states $\mathbf{S}$ are not observed. Thus, we are trying to maximise the incomplete data log-likelihood, $\log \mathcal{L}(\lambda;\mathbf{O})$. The incomplete data log-likelihood can then be written as:

$$\log \mathcal{L}(\lambda;\mathbf{O}) = \log P(\mathbf{O};\lambda) = \log \sum_{\mathbf{S}} P(\mathbf{S};\lambda) P(\mathbf{O}|\mathbf{S};\lambda) \tag{8.10}$$

Through Jensen's inequality, we arrive at an auxiliary function $\mathcal{Q}$ (Eq. 8.11;

see Appendix B.1 for derivation), which we maximise at each iteration with respect to the model parameters, λ (Eq. 8.12):

$$\mathcal{Q}(\lambda, \lambda^n) = E_{\mathbf{O},\lambda^n}[\log P(\mathbf{O}, \mathbf{S}; \lambda)] \tag{8.11}$$

$$\lambda^{n+1} = \underset{\lambda}{\operatorname{argmax}}\, \mathcal{Q}(\lambda, \lambda^n) \tag{8.12}$$

where λ^n are the current parameter estimates.

Note that in Eq. 8.11, $\mathbf{O}$ and λ^n are constants. λ are the parameters we wish to adjust, and $\mathbf{S}$ is a random variable governed by the conditional distribution $P(\mathbf{S}|\mathbf{O}; \lambda^n)$. At each iteration of Eq. 8.12, the log-likelihood is guaranteed to increase (Murphy, 2012).

8.3.5.2 Useful Variables

As is convention, we define the forward variable, $\alpha_t(i)$, as the joint probability of the partial observed locomotor sequence (up to time t), and the hidden phase state at t being q_i, for a particular set of model parameters λ:

$$\alpha_t(i) = p(O_1, O_2, ...O_t, S_t = q_i; \lambda) \tag{8.13}$$

Similarly, the backward variable, $\beta_t(i)$, is defined as the joint probability of the partial observed locomotor sequence ($t+1$ to T), given the hidden phase state at t is q_i, and again for the particular model parameters λ:

$$\beta_t(i) = p(O_{t+1}, O_{t+2}, ...O_T | S_t = q_i; \lambda) \tag{8.14}$$

We can in turn define the probability, $\gamma_t(i)$, of being in hidden phase state q_i at

time t, given the full observed locomotor sequence and the model parameters λ:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(\mathbf{O};\lambda)} \tag{8.15}$$

And the probability, $\xi_t(i,j)$, of a hidden phase path being in state q_i at time t, and making a transition to state q_j at time $t+1$, given the full observation sequence $\mathbf{O}$ and the model parameters λ:

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{p(\mathbf{O};\lambda)} \tag{8.16}$$

8.3.5.3 Standard Parameter Update Rules

Maximising the $\mathcal{Q}$ function, and using the variables defined in Section 8.3.1, we arrive at the parameter re-estimation formulas:

$$\bar{\pi}_i = \gamma_1(i), \qquad 1 \leq i \leq N \tag{8.17}$$

which is simply the probability of being in state q_i at time 1.

$$\bar{b}_j(k) = \sum_{\substack{t=1 \\ O_t=v_k}}^{T} \gamma_t(j) / \sum_{t=1}^{T} \gamma_t(j) \tag{8.18}$$

which is the ratio between the expected number of times of being in state j and observing activity level v_k, and the total expected number of times of being in state j. By learning the emission probabilities on free-running locomotor data, we can ignore ΔZ for the emission probability update rule.

8.3.5.4 New Learning Rules for κ and μ

In Section 8.3.2.2, I proposed an IOHMM architecture with circadian constraints on the class of possible transition probability structures. This required the derivation of new parameter update rules. Here, learning occurred, not for individual transition probabilities (as in the standard framework), but for the κ and μ hyper-parameters specified in Eq. 8.5.

Upon examination, no solution exists in closed form for the maximum likelihood estimate of these hyper-parameters. Consequently, 'generalised expectation maximisation' (GEM) was used, where gradient optimisation methods adjusted parameter values *in the direction of* maximising the $\mathcal{Q}$ function at each iteration (Dempster et al., 1977).

To accelerate the GEM learning process, a momentum method was also implemented in the gradient optimisation (Qian, 1999). This involved adding a fraction (0.9) of the parameter update vector from the previous iteration to the current update vector. Furthermore, a smoothness penalty term was introduced to the objective function to prevent large discrepancies between κ_i, κ_{i-1} and κ_{i+1}, and μ_i, μ_{i-1} and μ_{i+1}. This acted to encourage smooth variations in the dynamics of neighbouring phase states, reducing also the dimensionality of the fitting problem and the corresponding demand on training data. This penalty can be considered a prior over smooth circadian dynamics in the model.

A full derivation of the κ and μ update rules can be found in Appendix B.2. These are defined here as:

$$\begin{aligned}
\bar{\mu}_{i_{del}} &= \mu_{i_{del}} + \upsilon \frac{\partial \mathcal{Q}}{\mu_{i_{del}}} \\
\bar{\mu}_{i_{adv}} &= \mu_{i_{adv}} + \upsilon \frac{\partial \mathcal{Q}}{\mu_{i_{adv}}} \\
\bar{\kappa}_{i_{del}} &= \kappa_{i_{del}} + \upsilon \frac{\partial \mathcal{Q}}{\kappa_{i_{del}}} \\
\bar{\kappa}_{i_{adv}} &= \kappa_{i_{adv}} + \upsilon \frac{\partial \mathcal{Q}}{\kappa_{i_{adv}}}
\end{aligned} \tag{8.19}$$

where υ is the learning step size for the GEM optimisation.

8.4 Training Data Simulations

The process of constructing a probabilistic model, such as an IOHMM, often requires data on which to 'train' the parameters of the model (see discussion of parameter learning in Section 8.3.5). However, in order to validate the performance of this trained model at inferring hidden clock phase, it stands to reason that we must ourselves know the correct answer − such that we can compare model predictions against an objective truth. In these situations, it can be useful to construct a completely separate model to simulate training data. This *in silico* data can be designed in such a way as to possess similar dynamics to the biological system we seek to model. With explicit knowledge of the data-simulating process, the researcher gains a benchmark against which to compare the predictions of the probabilistic model. For this reason, I used the previously-developed dynamical system model described in Section 8.2 as a separate data-simulating model to train my IOHMM.

8.4.1 A Discrete Dynamical System Model of Clock Phase

As described previously in Eq. 8.2 (page 126), a linear discrete dynamical system model of clock phase was constructed, where phase step-size, w, was chosen to match the time resolution of the experimental locomotor data, which is measured as the number of IR beam breaks per 5 min (see Section 4.1.1).

The functional form of the PRC term was modelled on the shape of experimentally-derived light PRCs discussed in Chapter 2 (Leloup and Goldbeter 1998; Stanewsky et al. 1998; Vinayak et al. 2013; page 52). This was represented functionally using a fifth-order Fourier series (Fig. 8.7; see Appendix B.3 for details).

Phase adjustments of the dynamical system model to a perturbing Zeitgeber stimulus was therefore determined by the point in the cycle at which the stimulus was administered; sometimes advancing, sometimes delaying, and sometimes showing no response at all. Fig. 8.8 C illustrates entrainment of this simulated clock phase to a periodic Zeitgeber signal.

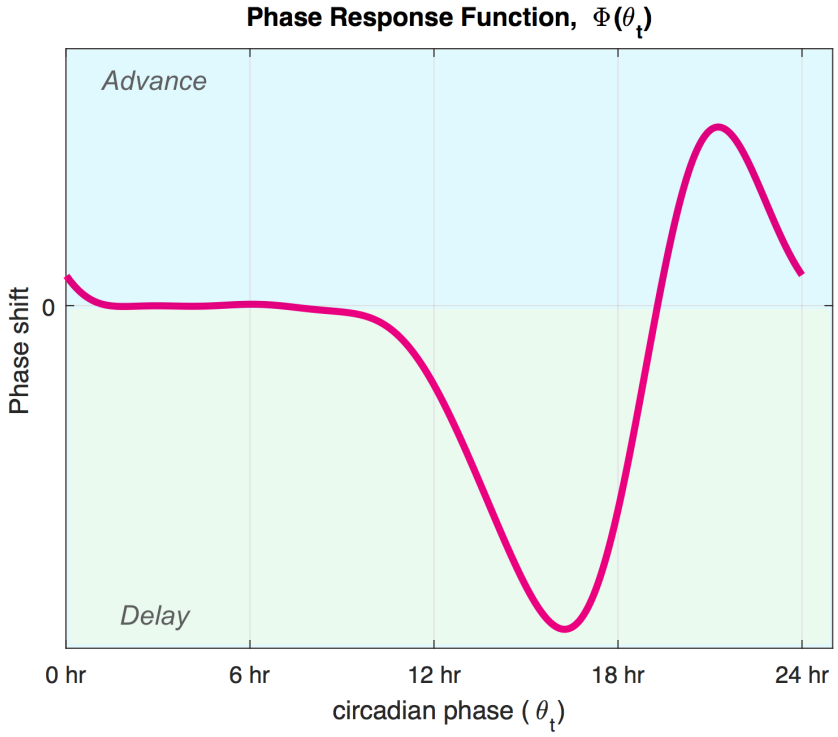

Figure 8.7: Functional form of the PRC term, $\phi(\theta_t)$, included in the data-simulating dynamical system model described in Eq. 8.1. A fifth order Fourier series was fit to resemble experimental observations (Vinayak et al., 2013). Note this PRC specifies advance and delay responses of the clock phase to a Zeitgeber pulse, as well as the presence of a 'dead zone' where no phase change occurs, as described in Johnson (1999). Generated in *MATLAB*.

8.4.2 A Gaussian Process Model of the *Drosophila* Locomotor Waveform

To simulate realistic data, it was first necessary to construct a mathematical representation of the activity waveform, $f(\theta_t)$, given the phase of the clock (Eq. 8.2 on page 126). As described in Chapter 2, one possible option for modelling this waveform is parametrically, via a combination of sine and cosine functions. However, this involves the often cumbersome task of devising the precise functional form of the linear model to be used.

Gaussian processes (GPs) offer an alternative, ostensibly non-parametric, method to regress data. An excellent review of GPs can be found in Rasmussen and Williams (2005). Briefly, the GP is a generalisation of the Gaussian distribution, which is completely defined by its mean and covariance. Much as sampling from a

2-dimensional Gaussian distribution generates a 2-dimensional data point, we can think of a collection of N data points in a time series as being one sample from an N-dimensional Gaussian. As N grows very large, this long data vector can be thought to approximate a function, with the covariance matrix of the N-dimensional distribution specifying the class of functions that are possible. In this way, GPs specify a probability distribution over functions, which can then be conditioned on observed data points according to Bayesian methods.

While GPs mitigate the need for an exact functional form of the data, they do make use of a covariance matrix-generating kernel, referred to as the 'covariance function', which is itself parametrised by a small number of 'hyperparameters'. It could therefore be argued that GPs are not completely devoid of parametric specification. However, in practice, the choice of covariance function is limited to a few common candidates − in this study, I opted to use the squared-exponential function, which can describe any smooth function:

$$k_{SE}(\theta_t, \theta_{t'}) = \sigma^2 \exp(-\frac{(\theta_t - \theta_{t'})^2}{2l^2}) + \delta_{tt'}\varepsilon^2 \tag{8.20}$$

where $k_{SE}(\theta_t, \theta_{t'})$ specifies the covariance between two independent variables in the multidimensional Gaussian distribution, σ^2 is the signal variance determining the average distance of the function away from its mean, l is the lengthscale parameter determining the distance between variables that co-vary (this governs how 'wiggly' the function can be), and ε^2 is a noise parameter for data around the function, which applies only when $\theta_t = \theta_{t'}$, as shown by the delta function, $\delta_{tt'}$.

A GP model was fit to one full day of experimentally-derived free running locomotor data (averaged across 160 wild type flies) by minimizing the negative log marginal likelihood with respect to the hyperparameters (Eq. 8.21; Fig. 8.8 A). This was able to simulate realistic activity recordings from the clock phase trace provided by Eq. 8.1.

$$f(\theta_t) \sim \mathcal{GP}(0, k_{SE}(\theta_t, \theta_{t'})) \tag{8.21}$$

Thus, combining the GP model with the phase output of Eq. 8.1, activity rhythms were simulated that displayed 24-hour free running period, and were entrainable to an external Zeitgeber through biologically plausible phase response dynamics (Fig. 8.8 B; Fig. 8.7). Importantly, the state of the underlying circadian phase for a given locomotor trace is known explicitly (Fig. 8.8 C; Eq. 8.1).

8.4.3 Simulating Masking Behaviour in the Training Data

To simulate training data that displays the masking characteristics we wish to infer using the IOHMM, an adjustment was made to the data-simulation model described in the previous section. Simulated activity was again generated according to a GP model (described in the previous section), as this is believed to capture the free running circadian activity waveform. Separately, as part of a *non-circadian* process, the simulated activity values were multiplied by a startle strength factor, *SS*, at times corresponding to $\Delta Z_t = 1$. To illustrate this, example activity traces for different *SS* magnitudes are shown in Fig. 8.9.

8.5 Results

8.5.1 IOHMM Captures the Statistical Structure of *In Silico* Free Running Data

The IOHMM was trained on 20 free running locomotor traces, each spanning 10 days at 30 min resolution (480 data points per trace), and simulated as described in Section 8.4 ($\forall t : Z_t = 0$). Model parameters were initialised uniformly as shown in Fig. 8.10, and learning was allowed to continue until either the log likelihood of the objective function varied by less than 10^{-6}, or the iteration number reached a maximum value of 5000.

Parameter learning on this free running data acted to increase μ values, such that the range of possible transitions in the hidden phase space was narrowed, and

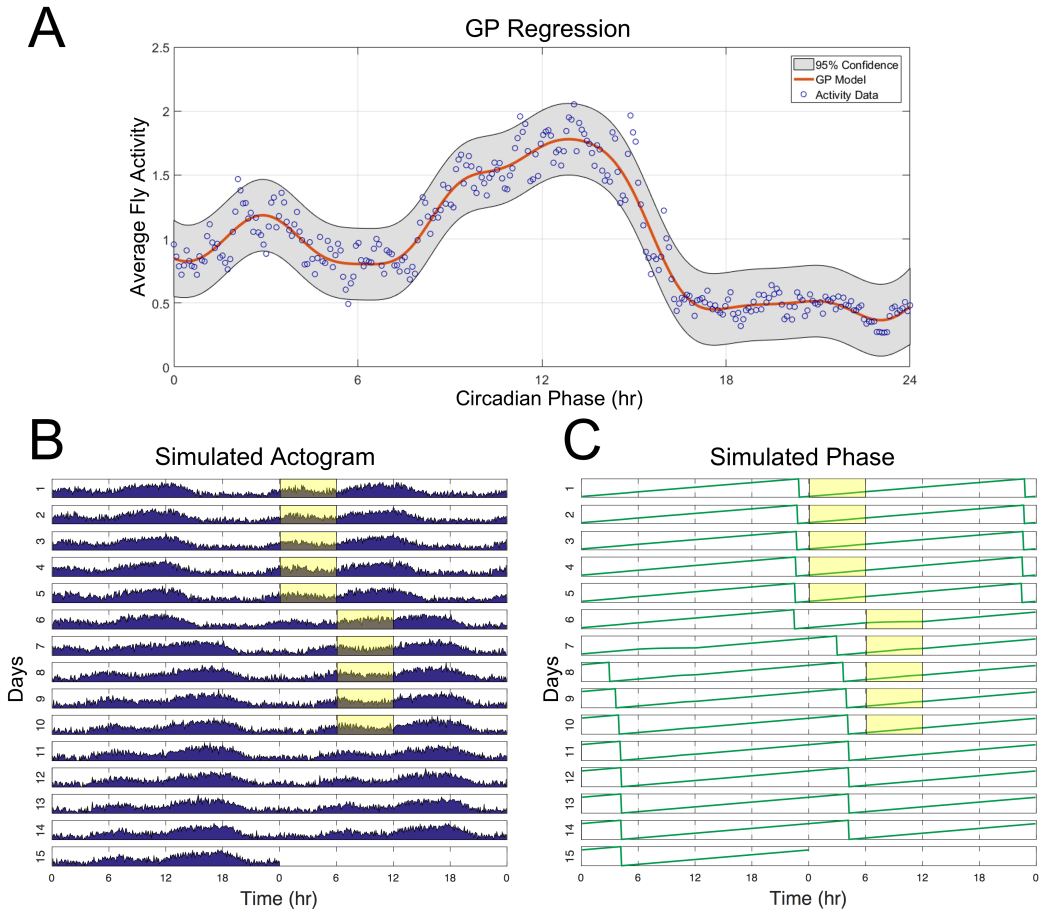

Figure 8.8: Simulations from the dynamical system model. (A) GP regression on 24 hours of wild type locomotor data (blue circles) averaged over 160 flies. A squared-exponential covariance function was used (Eq. 8.20). Orange line shows the most likely function sampled from the fitted GP distribution, and grey shaded area shows the 95% confidence bounds on this distribution. Fitted parameters were $\sigma^2 = 0.35$, $l = 1.55$, and $\varepsilon^2 = 0.136$. (B) Simulated locomotor activity from the phase trace described in (C) during entrainment to 6 hr pulses of an external Zeitgeber stimulus (yellow rectangle). Note the characteristic bimodal profile in the waveform. (C) Phase trace simulated from the model described in Eq. 8.1. The PRC term of the model allows the phase to entrain to the Zeitgeber regime, which goes on to simulate locomotor behaviour in (B) through sampling from the model in (A). Zeitgeber regime (yellow) only shown for the second half of the double actogram to make the simulated data more easily viewed.

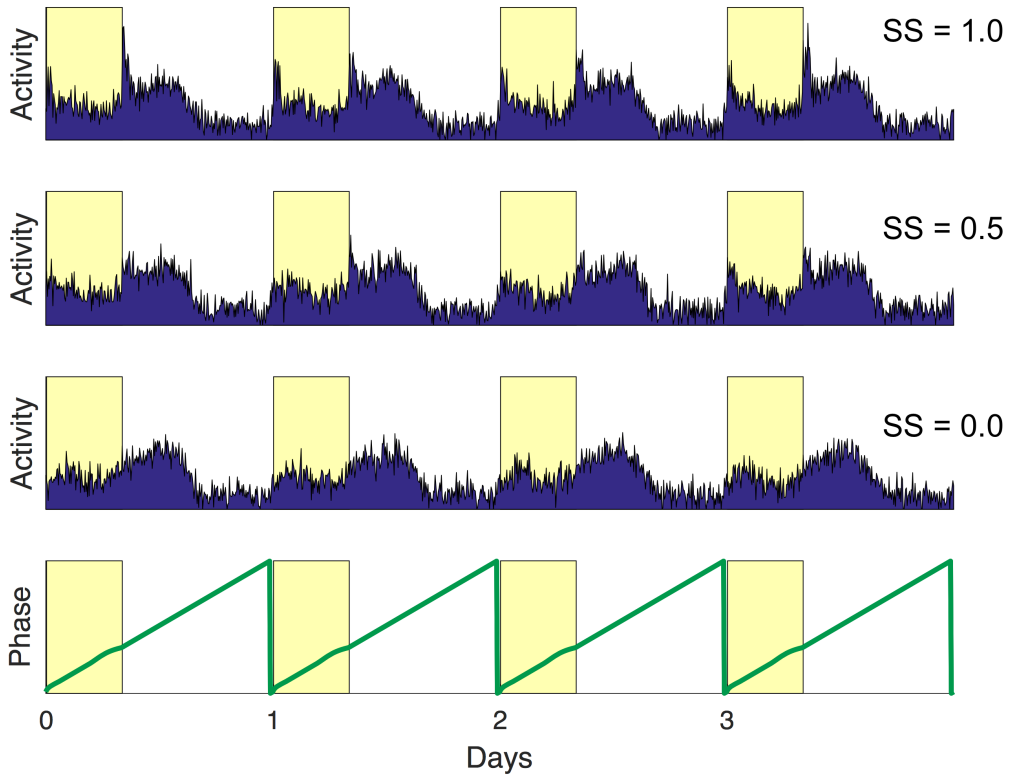

Figure 8.9: Training data simulations incorporating *Drosophila* startle behaviour. Simulated locomotor activity from a single phase trace (bottom, green) during entrainment to 8 hr:16 hr Zeitgeber cycles (yellow rectangles show presence of Zeitgeber). Locomotor traces in the top three panels were simulated using different startle strengths ($SS = 0$, 0.5, and 1.0). Note that as SS increases, there is a more pronounced peak of activity (or startle) at the points of Zeitgeber onset/offset.

the model approximated a left-right architecture (Fig. 8.10). This agrees with the notion of a free running clock where phase is unperturbed and evolves sequentially in time. Note that for the free running case, $Z_t = 0$, and thus κ exerts no effect on the transition probabilities (see Eq. 8.5 on page 135, and Eq. 8.6 on page 136). Accordingly, no learning occurs for these model parameters.

The emission probability matrix, B, was fit to capture the functional form of the locomotor activity. As one might have hoped, these learned emission probabilities reveal striking similarities with the functional form of the activity defined by the dynamical system model (Fig. 8.8). Clearly, the IOHMM is able to learn the statistical structure of the simulated activity data.

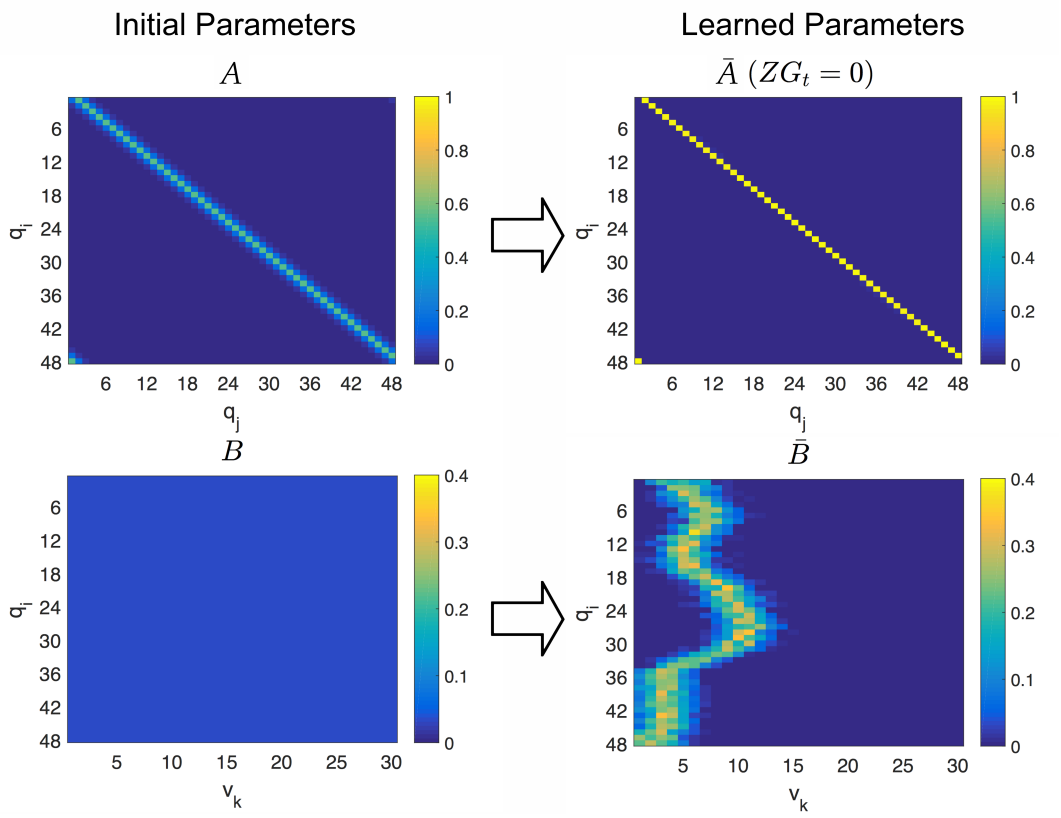

Figure 8.10: IOHMM parameter learning on FR training data. Training of the transition probability kernel, k_A, hyper-parameters, μ and κ (top), and the emission probabilities, $B(0)$ (bottom). Initial parameters were $\mu_{i_{init}} = 0.84$ (for all i), $\kappa_{i_{init}} = 0$ (for all i), and $\eta_{init} = 0$ (top left); and $b_{ik}(0) = \frac{1}{30}$ (bottom left). Note that μ and κ define τ (Eq. 8.6), which is used to evaluate the transition probability matrix (Eq. 8.5). During free run, $Z_t \neq 1$, and thus κ does not contribute to $A(Z)$. The learned transition probability structure for the IOHMM approximates a left-right architecture (top right), and the emission probability structure captures the locomotor waveform across a single day (bottom right).

8.5.2 IOHMM Learning Captures Circadian Phase Responses to a Zeitgeber

A new training dataset was generated as described in Section 8.4, with $SS = 0$, such that the data displayed no startle behaviour. 20 locomotor traces were simulated from a clock that was sporadically pulsed by a 5 min Zeitgeber signal (once per day) to perturb the underlying phase sequence (for an example, see Fig. 8.11 A). Learning was performed for the κ and η parameters only, which were initialised as before: $\kappa_{i_{init}} = 0$ (for all i) and $\eta_{init} = 0$. The μ, π and B parameters were set to the values learned through training on free running data (Fig. 8.10 right).

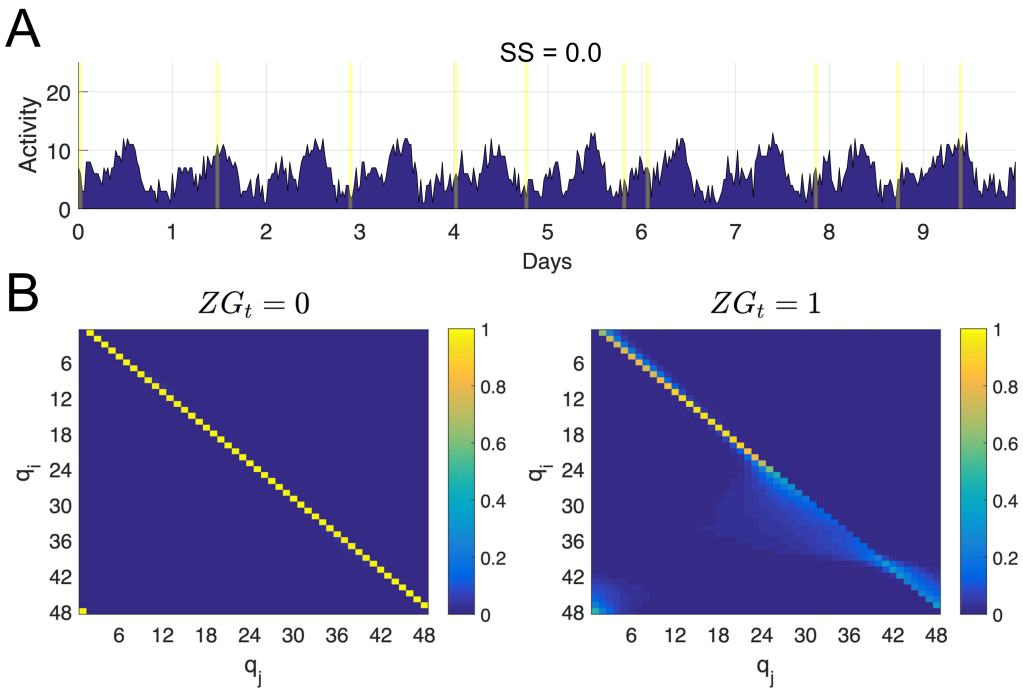

Figure 8.11: IOHMM learning of the transition probability structure in perturbed locomotor data. (A) Example training data trace. 10 days of locomotor activity simulated by a dynamical system model of the clock. One 5 min Zeitgeber pulse was administered each day (yellow bars), and *SS* was set to 0 such that no additional startle behaviour was present in the data. (B) Learned transition probability matrices. The κ parameters were fit using GEM using gradient optimisation methods. When $Z_t = 0$ (left), the IOHMM determines clock phase is likely to evolve sequentially in time; when $Z_t = 1$ (right), the system displays richer transitional dynamics.

Accordingly, two transition probability structures were learned - one describing phase dynamics in the absence of a Zeitgeber (governed exclusively by μ), and one in the presence of a Zeitgeber (captured by κ). For $Z_t = 0$, the trained IOHMM approximated a left-right architecture as expected.

For $Z_t = 1$, the learned κ values describe more complex hidden clock phase statistics, permitting larger jumps in the phase space. The structure of these clock phase statistics bear a strong resemblance to the PRC term that was explicitly programmed into the data-simulating model (Fig. 8.7; Fig. 8.12).

This is an important point − through analysis of locomotor traces lasting only 10 days, the IOHMM was able to recover the phase response dynamics of the circadian system that generated the behaviour. The ability to learn these parameters

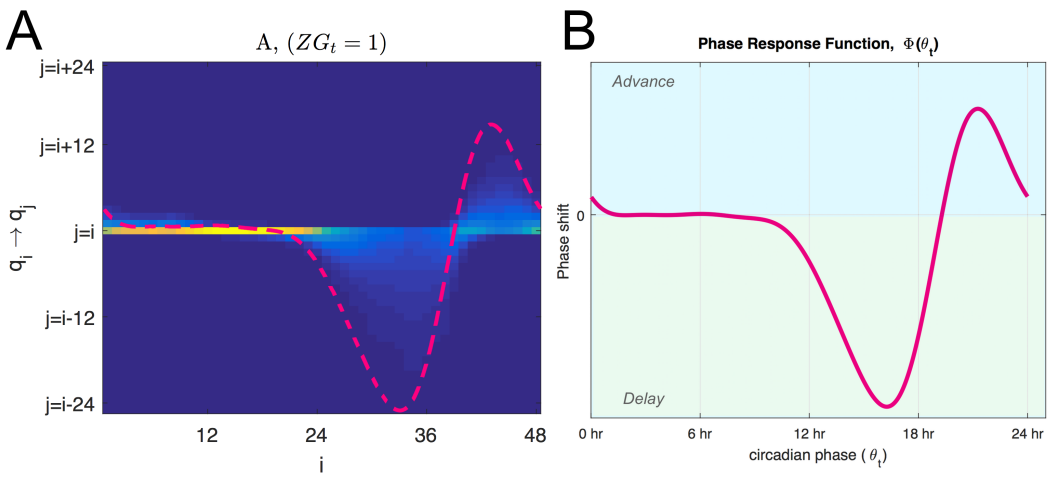

Figure 8.12: Training of the IOHMM recovers the true PRC. (A) The learned transition probability matrix from Fig. 8.11 B ($Z_t = 1$) reformatted such that the transition probabilities for q_i to q_j for $i = j$ lie horizontally along the mid line. This is equivalent to a phase shift of 0 hr. The functional form of the PRC shown in (B) is overlaid in the pink dashed line. (B) The PRC term programmed into the data-simulating dynamical system model of the circadian clock (taken from Fig. 8.7). This term governs the phase response dynamics of the simulated locomotor data, which were used to train the IOHMM. κ parameter learning in the IOHMM is able to correctly recover these dynamics, yielding tremendous utility for the study of other circadian systems.

from a small training dataset is likely to be due in part to the smoothness criteria enforced on page 144

8.5.3 IOHMM Learning Captures Masking Behaviour

Five separate IOHMMs were trained, each on a new dataset generated as described in Section 8.4 (adapted on page 148), and differing only in the degree of startle displayed ($SS = 0$, 0.25, 0.5, 0.75 and 1.0). As before, each training dataset comprised 20 locomotor traces simulated from a clock that was sporadically pulsed by a 5 min Zeitgeber signal (once per day) to perturb the underlying phase sequence (see Fig. 8.13 A for examples).

Regardless of the magnitude of startle displayed in the training data, all five IOHMMs learned similar values for κ (Fig. 8.13), all of which captured the PRC form defined in Fig. 8.7. Instead, differences in the training data subsets were captured through the correct learning of η, with larger η ascribed to data generated using larger SS (Fig. 8.13 D).

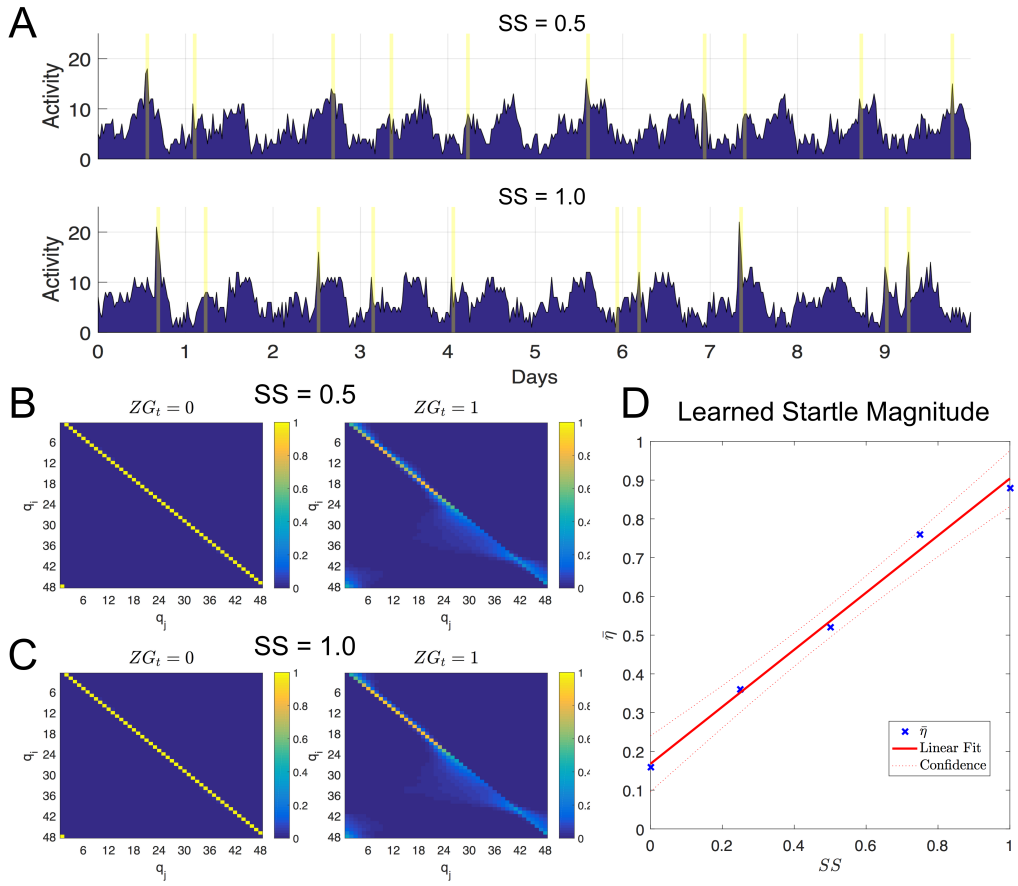

Figure 8.13: IOHMM learning of the transition probability structure in perturbed locomotor data with startle behaviour. (A) Example training data traces. 10 days of locomotor activity simulated by a dynamical system model of the clock. One 5 min Zeitgeber pulse was simulated each day (yellow bars), and SS was set to 0.5 (Top) and 1.0 (Bottom). (B,C) Learned transition probability matrices for training data where $SS = 0.5$ (C) and $SS = 1.0$ (B). The κ parameters were fit using GEM and gradient optimisation methods. When $Z_t = 0$ (left), the IOHMM determines clock phase is likely to evolve sequentially in time; when $Z_t = 1$ (right), the system displays richer transitional dynamics. (D) Learned startle magnitude η for training data displaying either $SS = 0, 0.25, 0.5, 0.75, 1.0$. The fitted η parameter increases linearly (red line) to capture the increased startle simulated in the training data.

8.5.4 IOHMM Clock Phase Inference on Simulated Locomotor Data

To evaluate performance of the IOHMM at inferring circadian phase, five new 'test' datasets, each comprising 20 10-day locomotor traces, were generated in the same manner as the training data described in the previous sub-section. It is important to note that these test locomotor traces were not presented to the HMM prior to testing (i.e. they were entirely separate from the data used to train the model).

Five separate IOHMMs were trained as described in Section 153. Given the observed locomotor traces and the learned model parameters, the Viterbi algorithm was used to infer the most likely hidden clock phase sequence that generated the relevant test data. A performance metric, $\mathcal{J}$, was defined as the summed absolute deviation of model predictions from true phase values, divided by the predictions of a naive model that simply predicted constant phase at the midway point of 12 hr (see Appendix B.4; Fig. B.1). Here, a score of 1 denotes equivalent performance to the naive model, and a score of 0 denotes perfect inference.

All IOHMM models performed sensible inference, with hidden clock phase predictions closely following the true phase of the simulated clock driving the data traces (Fig. 8.14 blue). To further evaluate the benefits of the input-output structure, each of the five models was again trained on the same dataset, however this time omitting information about the Zeitgeber signal. These models, which I refer to simply as HMMs due to the fact that they receive no input, were able to perform clock phase inference on the locomotor data (Fig. 8.14 red). However, as expected, they produced more incorrect predictions at the points of Zeitgeber administration. This is also shown by the consistently larger $\mathcal{J}$ scores for the HMMs compared to the IOHMMs (Fig. 8.15). Together, these data show that the specially-designed IOHMM is an improvement on the standard HMM framework for modelling the *Drosophila* circadian clock.

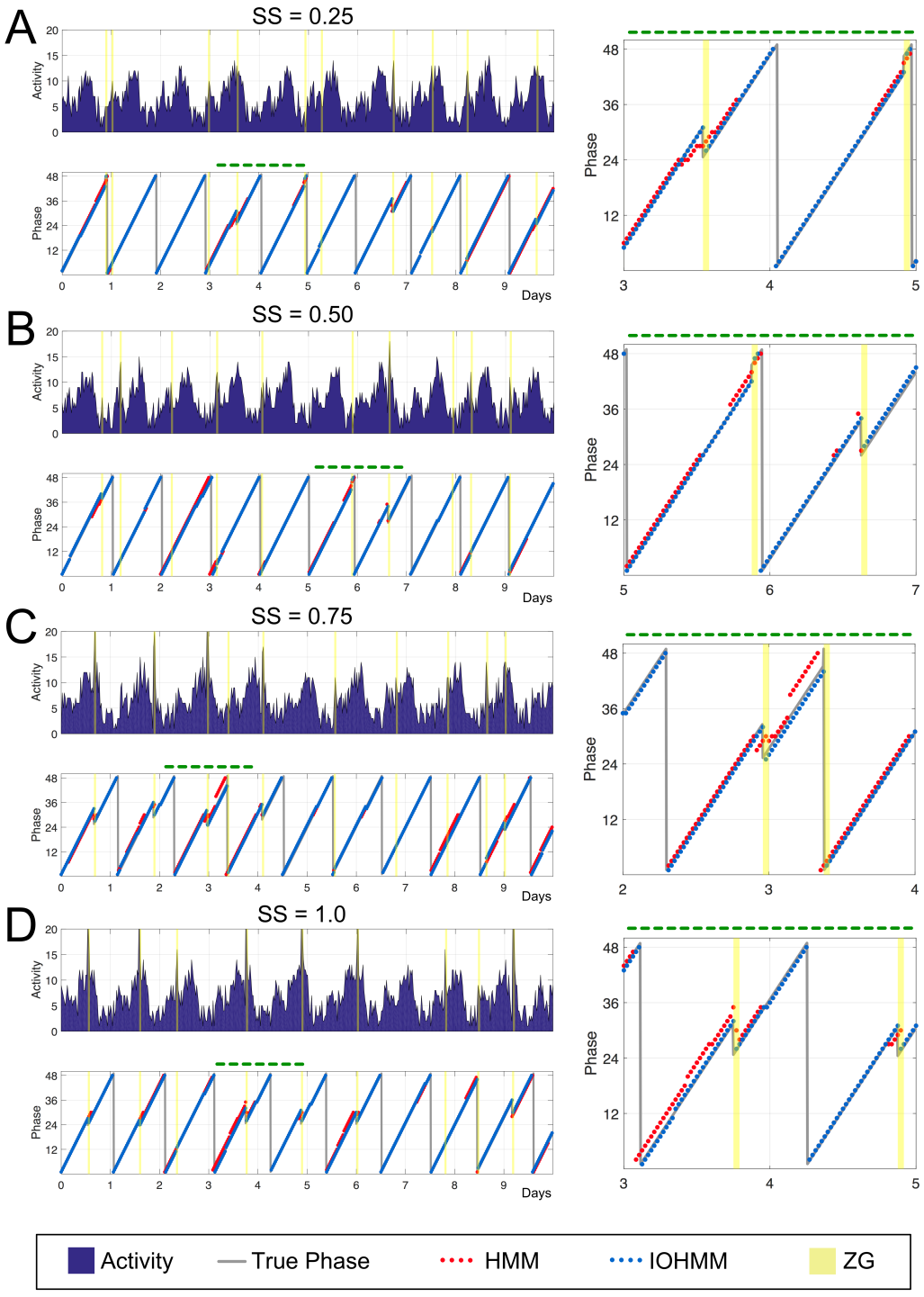

Figure 8.14: IOHMM and HMM phase inference on Zeitgeber-pulsed data displaying behavioural masking. Predictions of an IOHMM and an HMM trained on locomotor data generated from an *in silico* clock that was repeatedly perturbed by a Zeitgeber stimulus (yellow bars), displaying different magnitudes of startle behaviour. Model predictions were evaluated against previously unseen 'test' datasets generated in the same way as the training data. Training/test data displayed differing startle magnitudes: $SS = 0.25$ (A), $SS = 0.5$ (B), $SS = 0.75$ (C), $SS = 1.0$ (D). Note the improved performance of the IOHMM over the HMM, particularly at points where the Zeitgeber stimulus is administered, highlighted in the magnified plots to the right (delimited by green dashed lines).

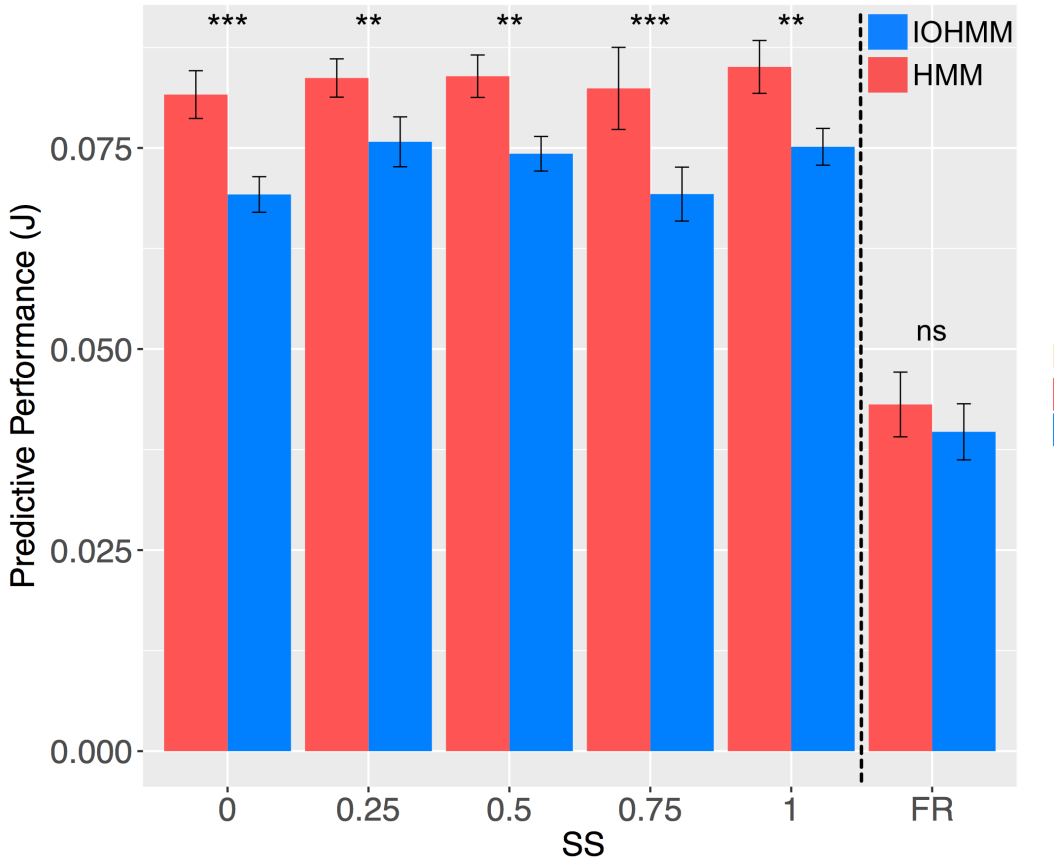

Figure 8.15: Comparison of clock phase inference between an IOHMM and an HMM. Models were trained on locomotor data generated from an *in silico* clock that was repeatedly perturbed by a Zeitgeber stimulus, displaying different magnitudes of startle behaviour (see Fig. 8.14). $\mathcal{J}$ values show the discrepancy between model predictions and true phase values for each test trace, normalised against the predictions of a naive model, for both an HMM (red) and IOHMM (blue). Performance on free running data is shown on the right, separated by a black dashed line. The IOHMM consistently performs better than the HMM on data generated in the Zeitgeber condition, as shown by the lower $\mathcal{J}$ values. ($p < 0.05*, p < 0.01**, p < 0.001***$). Error bars show SEM.

8.6 Analysis of a Biological Dataset

Up to this point, my statistical approach to modelling *Drosophila* circadian locomotor rhythms has focussed on *in silico* data generated by the discrete dynamical system model described on page 145. A key rationale behind this approach was that explicit knowledge of the *in silico* data mechanics provides a means to validate predictions of the HMMs and IOHMMs. Having demonstrated the promise of these statistical models on simulated circadian data (Fig. 8.15), I next implemented an IOHMM for the analysis of 'real-world' locomotor data obtained in wild type flies.

8.6.1 Materials & Methods

8.6.1.1 Fly Strains

Three wild type strains were used: Canton S, Oregon R and white-eyed (w^{1118}) flies, all obtained from Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu). Flies were reared under 12:12 LD cycles on *Drosophila* medium (0.8% agar, 2.2% sugar-beet syrup, 8.0% malt extract, 1.8% yeast, 1.0% soy flour, 8.0% corn flour, and 0.3% hydroxybenzoic acid) at 25°C and 60% humidity. Only male flies at an age of 3 to 6 days were used in experiments.

8.6.1.2 Activity Monitoring

Locomotor activity rhythms were recorded automatically using the *Drosophila* Activity Monitoring (DAM) system (*Trikinetics*, Waltham, MA) as previously described in Chapters 1 and 4. Analysis of locomotor data was conducted as described in Chapter 4.

8.6.1.3 Environmental Regime

I have so far been conscious to maintain a Zeitgeber-agnostic approach to the work presented in this chapter. This has been to highlight the fact that the HMM framework can indeed be applied to the investigation of many different Zeitgeber stimuli. However, experimental replication of the *in silico* training results necessarily required a choice for the given Zeitgeber stimulus. For this, a 5 min light pulse was used to match similar perturbations applied in previous experiments investigating *Drosophila* locomotor PRCs (Stanewsky et al. 1998, Leloup and Goldbeter 1998,

Fig. 2.6).

Wild type flies were first entrained to three days of LD at 25°C, before entering 2 days of DD to establish endogenous locomotor rhythms. Light pulses were then administered sporadically for a number of days at intervals no smaller that 24 hr apart, so as not to induce too much masking behaviour, which would disguise the underlying rhythms in the data (Fig. 8.16).

Sequential perturbation of the underlying clock made it impossible to ensure that subsequent Zeitgeber stimuli were administered at a range of different clock phases. Put simply, circadian time becomes increasingly unclear after the first Zeitgeber pulse. In an attempt to mitigate the impact of Zeitgeber pulse redundancies, the dynamical system model described on page 145 was used to devise two experimental regimes where pulses were administered to cover a range of circadian times, accounting for shifts in the underlying clock phase (Fig. 8.16).

8.6.2 Wild Type Flies Remain Healthy During Light Pulses

To assess viability of wild type flies during the light pulse regime described described in Fig. 8.16, a preliminary locomotor assay was conducted using three separate wild type genotypes: Canton S, Oregon R, and w^{1118}. 32 flies of each genotype were entrained to 3 days of LD at 25°C. The flies then either entered the light pulse regime defined in Fig. 8.16, or constant darkness at 25°C. Comparisons between the two conditions were made for each genotype (Fig. 8.17).

Viability of the flies was quantified as the proportion of surviving flies on day 20, relative to the number on day 1. For light pulse versus DD conditions, percentage survival was 65.6% versus 65.6% in Canton S; 37.5% versus 9.4% in Oregon R; and 46.9% versus 31.3% in w^{1118} flies. Surprisingly, across two out of the three wild type lines, viability was higher during the light pulse condition than in DD. This provides evidence to suggest that the light pulse regime in Fig. 8.16 does not negatively affect fly health.

These results show no serious experimental issues associated with the process of generating biological training data for the IOHMM developed in this chapter. While Canton S flies displayed the highest survival rate across the two environmen-

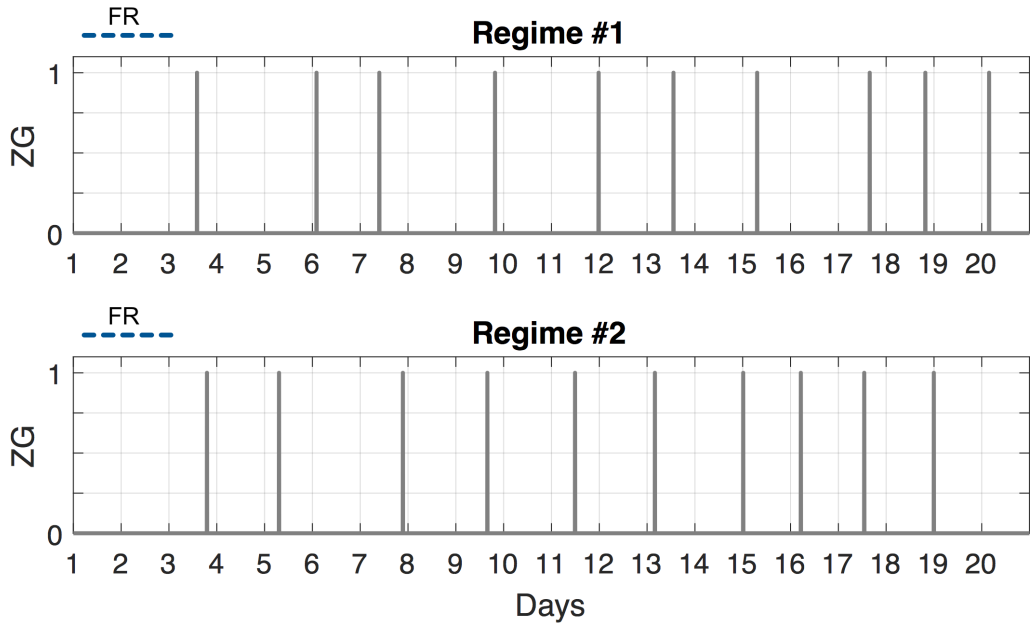

Figure 8.16: Light pulse regimes. Two schedules were devised for the administration of 5 min light pulses. Flies were first entrained to 12 hr: 12 hr LD at 25°C (not shown) before entering two days of free run (DD at 25°C) shown in the first two days of the Zeitgeber traces. After 2 days in free run, flies were sporadically pulsed by 5 min light stimuli, otherwise remaining in DD at 25°C. Pulses were designed to occur no closer than 24 hr apart, and a dynamical system model of clock phase was used to guide pulse administration in an attempt to cover a range of circadian times. Two regimes were used to match the number of incubators available for this experiment.

tal conditions, the w^{1118} genotype consistently displayed greater activity levels, as well as clearer behavioural rhythms, and a pronounced bimodal activity profile (Fig. 8.17). For these reasons, coupled with the fair health of this line, w^{1118} flies were chosen for subsequent experiments.

8.6.3 Training Data

Locomotor data was collected using the *Trikinetics* system as described in Materials & Methods (page 158). Using functions from the *Luceo* toolbox developed in Chapter 4, individual locomotor traces were cropped to exclude data collected after any point of fly death (as judged by activity levels falling consistently to zero).

As the *in silico* data was generated to emulate mean population activity of wild type flies (see Section 8.4.2), it was important to implement similar averaging of the biological data. Taking the mean across all flies, however, would generate only

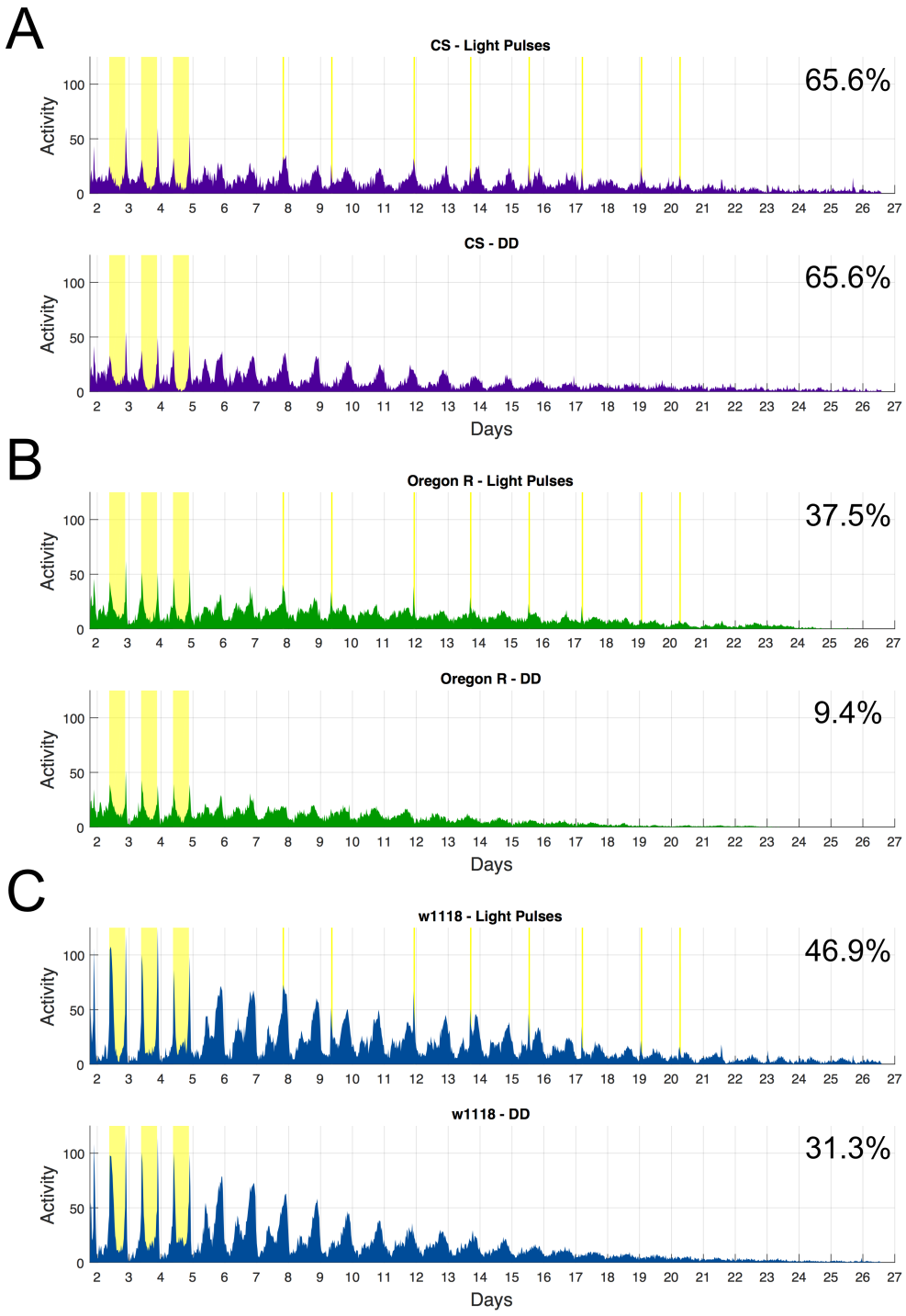

Figure 8.17: Wild type locomotor activity during DD or light pulse regime. Locomotor traces of either Canton S (A), Oregon R (B), or w^{1118} (C) during light pulse regime #2 from Fig. 8.16 (top) or free running DD at 25°C (bottom). Percentage survival of each group was calculated on day 20 of the experiment (displayed in the top right corner of the plots and reported in the text). Presence of the light Zeitgeber is shown in yellow. DAM2 monitors were used for data collection, due to equipment limitations, and the lack of a strong need for high resolution data in these experiments.

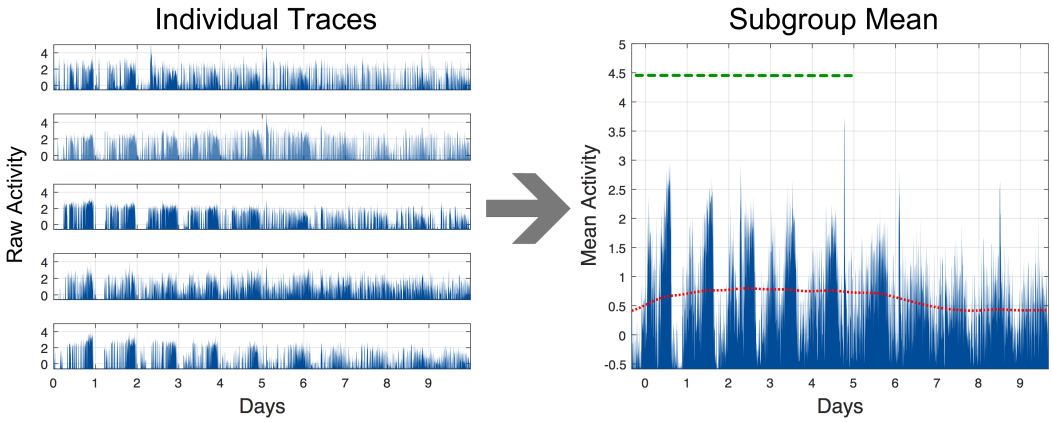

Figure 8.18: Preprocessing of training data. Individual traces were sorted according to length and grouped into subsets of five. A mean trace was calculated for each subset to produce 29 training examples for the IOHMM (15 for regime #1 and 14 for regime #2). Each mean trace was then de-trended as described previously (red dashed trend line illustrated in the right plot; see Chapter 4). Due to long-term changes in the activity waveform, only the first 5 days of data following the initial 3 days of entrainment to LD were used (green dashed line).

one training example for the IOHMM. To create a number of different training locomotor traces, data from individual flies were sorted according to length of the traces. These were then grouped into sets of five, and a mean trace was calculated for each group (length specified by the shortest trace in the subset), thus producing multiple average activity traces with which to train the IOHMM (Fig. 8.18). This process reduced the noise associated with individual locomotor traces, while preserving a diverse number of training examples. Each new training example was then de-trended and normalised as described in Chapter 4 (page 59).

The data presented in Fig. 8.17 displays a clear decrease in activity levels over time. Indeed, this is an expected feature of *Drosophila* locomotor behaviour during constant conditions (see discussion in Chapter 1, page 30). This decay was not completely removed by de-trending the data (Fig. 8.18). Moreover, interrogation of Fig. 8.17 also reveals long-term changes in the overall shape of the locomotor profile. Comparing the activity in the first few days of free run with that in later days shows a pronounced loss of the morning activity peak.

This is likely to present a challenge to the IOHMM, which, in its current form,

attempts to capture the statistical structure of the activity, independent of free run duration. Training on locomotor traces of the kind presented in Fig. 8.17 would therefore act to 'smear' the learned IOHMM emission probabilities, reducing the power of the model. For this reason, only the first 5 days of data after initial 12 hr:12 hr LD entrainment was used to train the IOHMM. This provided less variability in the statistics of the training examples, at the expense of decreasing the overall amount of data available. Note that this means only a limited number of light pulse events were explored (as many as occurred in the 3 days after free run; Fig. 8.16). The full utility of the experimental data collected, therefore, awaits further developments to the IOHMM (see Discussion).

8.6.4 An IOHMM Learns the Statistical Properties of *Drosophila* Locomotor Data

An IOHMM was trained on the experimental locomotor data, which comprised 24 locomotor traces. To learn the probability structure of the free running activity, the IOHMM was initially trained using the first two days of data (where no light pulse was administered; Fig. 8.16). Parameters were initialised uniformly to match those used for the *in silico* experiments discussed previously: $\mu_{i_{init}} = 0.84$ (for all i), $\kappa_{i_{init}} = 0$ (for all i), B was uniformly populated ($1/30$), and $\eta_{init} = 0$.

The emission probability matrix, B, was fit to capture the functional form of the locomotor activity (Fig. 8.19 A). As observed for the simulated training data, training on biological data resulted in an increase in the magnitude of the μ parameters, such that the range of possible transitions in the hidden phase space was narrowed, and the model approximated a left-right architecture (Fig. 8.19 B).

The IOHMM was then re-trained using the first 5 days of light pulse regime # 1 (15 traces) or regime #2 (14 traces). Learning here was performed for the κ and η parameters only, which were initialised as before: $\kappa_{i_{init}} = 0$ (for all i) and $\eta_{init} = 0$. The μ, π and B parameters were fixed to the values learned through training on free running data (Fig. 8.19 A,B).

On these training data, the IOHMM fit κ to learn a more flexible transition probability structure, describing responses of clock phase to the presence of the light

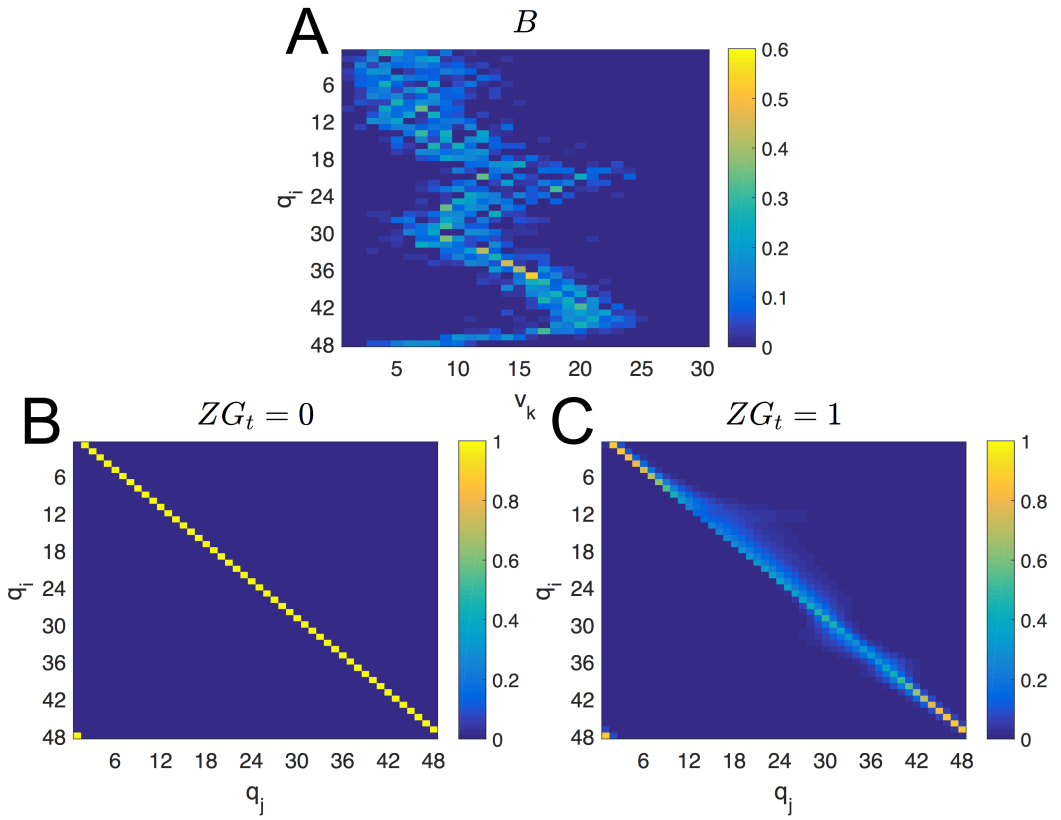

Figure 8.19: IOHMM parameter learning on *Drosophila* locomotor data. Learning of the emission probability structure (A), and the transition probability structure in the absence (B) and presence of (C) the light Zeitgeber. (A) was trained on the first two free running days of the training data to capture the intrinsic locomotor waveform. The κ parameters were fit with GEM and using gradient optimisation methods on the full five days of training data (C). When $Z_t = 0$ (B), transition probabilities are defined by μ, and the model determines clock phase is extremely likely to evolve step-wise in time. When $Z_t = 1$ (C), the learned κ parameters define richer phase dynamics underpinning the experimental locomotor data.

pulse stimulus (Fig. 8.19 C). Inspection of these transition probabilities suggests the fly clock permits large jumps forward in the phase space around hidden states 10 to 24 (corresponding to 05:00 and 12:00 in real time). The 9AM:9PM LD regime used in the first three days of entrainment defines these absolute time values to be CT 20 and 03 respectively, which coincides with the advance regions in previously reported experimental light PRCs (Leloup and Goldbeter 1998; Stanewsky et al. 1998; Vinayak et al. 2013; Fig. 2.6, page 52). While a similar delay component of the experimental light PRC is not as obvious in Fig. 8.19 C, this likely reflects the limited number of Zeitgeber pulses used ($\times 3$), which may not have been sufficient

to fully probe the range of circadian dynamics underlying fly locomotor behaviour.

Curiously, learning of the η parameter did not alter its value from zero. Visual inspection of the training data supports the existence of startle behaviour (Fig. 8.17), suggesting that modelling startle as we have may not sufficiently explain this behaviour in the real locomotor behaviour (see Discussion).

8.6.5 IOHMM Phase Inference on *Drosophila* Locomotor Data

Owing to the limited amount of biological data, and the lack of a suitable validation method, given that the true clock phase in this data is not known *ad hoc*, phase inference was performed on the training data (rather than a 'test' set as used in the *in silico* experiments). This is not a problem, as we are not validating model performance against some known objective truth.

The Viterbi algorithm was used to infer the most likely hidden clock phase sequence responsible for generating each training locomotor trace. As all the data collected during a given light pulse regime are expected to have the same or similar hidden clock phase, IOHMM predictions for these two experimental groups were averaged and plotted (Fig. 8.20). The model appears to perform sensible inference on the biological data, showing little variability in predictions across locomotor traces (Fig. 8.20, light blue shaded area), and demonstrating phase jumps in response to the light pulses.

8.7 Discussion and Future Work

Hidden Markov models have contributed to many diverse areas of science, including speech recognition (Rabiner, 1989), computational genomics (Eddy, 2004), and decoding neural spike data (Escola et al., 2011). In this chapter, I have presented one of the first applications of this established ML approach to chronobiology. Through adjustment of the standard HMM framework, I designed a probabilistic model of the fruit fly clock, which, when trained on both *in silico* and *in vivo* data, was able to learn the statistical structure of circadian systems.

In Section 8.4, I used a separate data-simulating model, based on traditional dynamical systems used in the field, to create *in silico* datasets where the under-

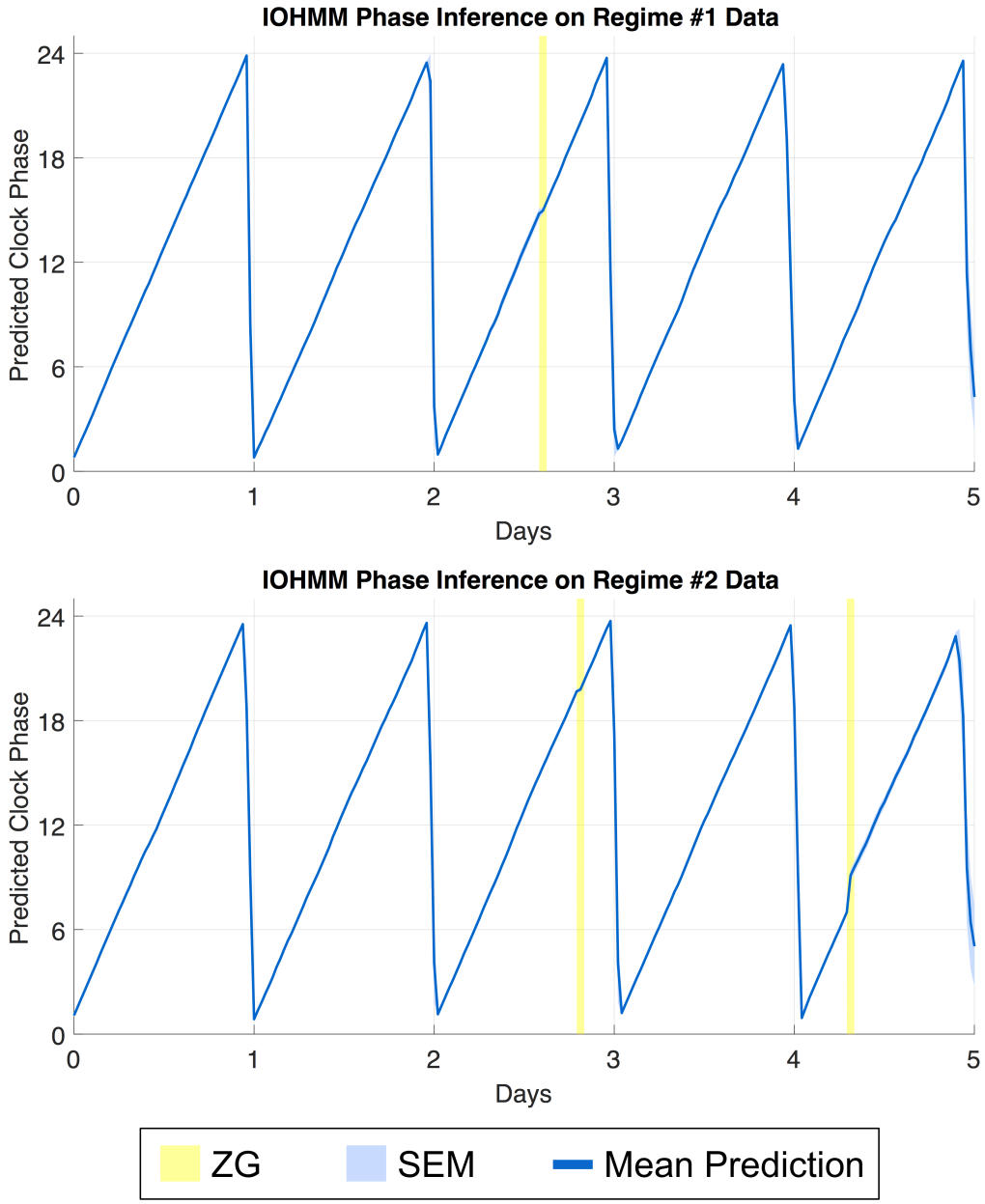

Figure 8.20: Clock phase inference of the IOHMM on wild type locomotor activity during exposure to 5 min light pulses. Parameters learned in training (Fig. 8.19) were used with the Viterbi algorithm to calculate the most likely hidden clock phase state sequence responsible for generating the experimental locomotor data. Plots show the mean predicted state sequences for data collected in light pulse regime #1 (top, n = 15) and #2 (bottom, n = 14). Note here that 'n' refers to the number of training examples, each of which comprised average data from 5 flies, making the overall number flies used 75 and 70 respectively. Blue lines show the mean predicted state sequence, light blue shaded area shows standard error of the mean, and yellow bars show light pulses, which are bordered by DD conditions at 25°C.

lying circadian mechanics were explicitly known to the researcher. This objective truth was then used to validate predictive performance of the IOHMM, which to a high degree of accuracy inferred the underlying phase driving the locomotor traces analysed. Specifically, the input-output structure allowed the IOHMM to respond appropriately to an incoming Zeitgeber signal, improving inference compared to a version of the model that did not receive Zeitgeber information (here referred to as the HMM). This supports the notion that the bespoke circadian architectures developed in this chapter provides a more faithful representation of the clock, making it more suitable for the analysis of circadian data. As the IOHMM and HMM are distinct only when $Z_t = 1$, the discrepancy in performance between the two models is be expected to increase with the number and duration of the Zeitgeber pulses.

Further investigation of *in vivo* locomotor activity measured in wild type *Drosophila melanogaster* demonstrated that the IOHMM is capable of inferring clock phase from biological datasets. However, application of the model to longer time series was challenged by variations in the fly locomotor waveform during constant conditions. These included a pronounced decrease in activity levels, coupled with a gradual loss of the morning peak of activity, as shown by locomotor assays performed in three different wild type genotypes. Techniques for de-trending *luciferase* time series (see Chapter 4) did not fully remove the global decrease in locomotor activity, highlighting this as a true feature of circadian behaviour, requiring explicit modelling. Indeed, long-term changes to fly locomotor activity were expected (see Section 1.3.1), and can be seen throughout the literature, however rarely is this seemingly ubiquitous phenomenon the subject of rigorous investigation.

One approach to modelling these long-term variations in the locomotor waveform would be through the addition of a 'time' input signal to the IOHMM. This would then act to adjust the emission probabilities of the model in a time-dependent manner. The use of an emission matrix-generating kernel function (equivalent to the transition probability kernel in Eq. 8.5), would likely be necessary to reduce the number of parameters in the model, minimising the volume of data necessary for

learning these parameters. The challenge of such an approach would be devising a kernel that preserved the rich statistical structure of the activity waveform.

The process of learning IOHMM parameters itself yielded insight on the circadian system under investigation. Indeed, fitting the κ parameters on data driven by a sporadically perturbed underlying clock, captured the response dynamics of the system. For the *in silico* case, the functional form of the phase response curve was known *ad hoc* to be a fifth order Fourier series (Fig. 8.7; Appendix B.3). Learning IOHMM parameters using these simulated data recovered the true PRC, which was represented in the transition probability structure of the model (Fig. 8.12). This capability was robust to the degree of startle behaviour displayed by the data. Furthermore, fitting the IOHMM on *in vivo* data measured in wild type flies also led to variations in the transition probability structure, which resembled the known light PRC for *Drosophila* (Fig. 8.19 C; Vinayak et al. 2013). Curiously, the η parameter, which models startle behaviour, was not altered during fitting beyond the initial starting value of zero. This is counter-intuitive, given that the data appears to display startle in response to the light pulses. Further examination is required to explain this behaviour of the model. Future work might benefit from a richer description of this behavioural masking in the IOHMM.

The performance of the IOHMM on the experimental locomotor activity is promising, as only limited training data was used (29×5 day activity traces). This illustrates the potential of the IOHMM for analysing small datasets, and further hints at the benefits to come from accommodating longer time series in the model. Furthermore, a good performance on limited data suggests the IOHMM framework might be suitable for modelling individual flies, permitting a deeper level of analysis, and allowing for more precise investigations of sensory effects on the clock. Moreover, in the context of a potential proprioceptive pathway in the *Drosophila* circadian system (discussed in Chapter 7), analysis of individuals is paramount, due to the implication that subtle differences in the activity of each fly are translated to changes in clock phase. It would certainly be interesting to explore this hypothesis using the IOHMM framework. Such an investigation might require deviations

away from representing distinct activity levels in the model, and instead outputting the rate parameter of a Poisson distribution. This has the potential to better-capture the more binary characteristics of individual time series.

The experimental work performed in Section 8.6.3 implemented light as the Zeitgeber stimulus, owing to its historically well-established position in the circadian literature. However, my IOHMM is inherently agnostic to the Zeitgeber used, and could easily be extended to investigations of other sensory signals. Under the probabilistic framework developed in this chapter, it would be interesting to repeat similar experiments using regimes comprising temperature or vibrational pulses. Training an IOHMM in this way presents the possibility of capturing the statistical PRC for these stimuli, and indeed any entraining cue. Further development of the IOHMM structure to model multiple Zeitgeber input signals simultaneously could then yield deeper insight on the statistics of multisensory integration, for which the work presented in Chapters 5 and 6 would provide a worthy foundation.

In addition to modelling different input signals, my IOHMM is also inherently general in its output – for example, train it on *luciferase* reporter data, and it becomes a molecular model of the clock. This flexibility comes from fitting the emission probabilities, which could feasibly capture the statistical waveform of any circadian time series. Indeed, recall that the standard HMM developed by Bieler et al. (2014) was used to model YFP fluorescence in mammalian fibroblast cells. Thus, while my IOHMM has been prototyped in the context of *Drosophila* locomotor behaviour, it has far broader applications across the circadian field.

8.8 Summary

Recent years have seen the tremendous success of machine learning and statistical modelling across the academic disciplines. In this chapter, I developed a probabilistic model of the circadian clock, shaped by biological considerations. This input-output hidden Markov model was able to infer phase of the underlying clock from locomotor data, providing higher resolution phase analysis of circadian data than existing methods in the field. My model was able to capture the statistical

properties of both the locomotor waveform, and clock phase dynamics, for which the latter further revealed the statistical phase response properties of the system to repeated perturbations by a Zeitgeber. Together, these capabilities provide a novel tool for investigations of sensory input to circadian systems, particularly with regards to reducing resource demands of existing experimental approaches. In this way, I hope to have addressed objective 5 outlined in Chapter 3.

My model can be viewed either as an analysis tool, or a probabilistic representation of the circadian clock itself. As such, I hope to have illustrated the power of these approaches to explain, understand, and analyse biological systems. I eagerly anticipate further advances at the intersection between chronobiology and statistical modelling.

Chapter 9

General Conclusions

Throughout my PhD, I have taken a combined experimental and theoretical approach to studying the circadian clock in *Drosophila melanogaster*. While drawing from differing academic disciplines, my investigations can be unified by their focus on one key question: what exactly is the *task* of the circadian system? In my attempts to address this question, I hope to have contributed to perspective, as well as yielded mechanistic insight on general principles of clock function.

Circadian clocks must synchronise to periodic timing cues in their environments. However, the existence of more than one cue forces a problem of sensory integration. The clock must therefore extract some notion of the time of day from these many, potentially conflicting, signals. In this thesis, I challenged the fly clock, creating artificial misalignments between light and temperature cues in order to test the limits of the network. During intermediate Zeitgeber discrepancies, conventional locomotor rhythms in wild type flies broke down, giving way to a novel locomotor pattern I here refer to as 'Plateau' (P) behaviour (Chapter 5). This behavioural response to environmental discord was also associated with a collapse of molecular cycling in the central clock neurons of the fly brain (Chapter 5).

At all times, my experimental investigations were guided by theoretical considerations of the clock, which may explain why previous studies did not report my findings. Bayesian treatment of the problem specified the use of large-amplitude fluctuations in both Zeitgeber signals, with the aim of generating two reliable, yet discrepant, timing cues. It seems that when the 'tick' is a fair match for the 'tock',

inputs conspire to break the clock.

This jovial idea of a staccato ticking and tocking in the environment highlights another key facet of most laboratory-based investigations in chronobiology – the square-wave Zeitgeber profile, which is binary in being either ON or OFF. In order to derive temporal meaning from such a signal, the fly clock *must* integrate information over time; for which the hidden Markov model developed in Chapter 8 provides a statistical representation (evaluating probability of the true time of day given current, and past, observations).

As the field progresses toward more naturalistic studies of circadian systems, we can expect to better-encapsulate the animal's subjective experience, and thus devise a more faithful account of the multisensory challenges posed to the clock. A logical development on my research would be to explore these more natural phase discrepancies between sensory cues, using profiles modelled on realistic daily variations. I eagerly anticipate the results to come from such an approach.

Together, the data presented in Chapter 5 reveal the central clock network in *Drosophila melanogaster* is able to tolerate some, but not all, environmental conflicts. This raises similar questions of human circadian function, which shares many principles in common with that of the fruit fly (see Chapter 1; Helfrich-Förster 2004). In nature, many brain systems are subject to multisensory challenges. These may be similarly vulnerable to conflicting sensory input. My paradigm enables novel approaches to test cue integration in an experimentally tractable neuronal network model, which may be of particular appeal to computational neurobiologists investigating Bayesian mechanisms. Furthermore, I expect my work to be of immediate interest to biomedical researchers studying human clock function (and dysfunction) as they suggest a more concerted, and explicitly multisensory, approach to the circadian clock and its numerous associated disorders.

Beyond the brain clock network, circadian oscillators are present in tissues around the fly body. While the precise role of these peripheral clocks is less clear than their central counterpart, they too function within multisensory environments, and must solve the associated processing challenges therein. Building on the re-

sults of Chapter 5, I explored the response of these peripheral clocks to sensory conflict. While a P behavioural phenotype was observed during misaligned conditions, molecular rhythms were largely unchanged in their waveform, and instead preferentially synchronised to the light stimulus (Chapter 6). These data reveal yet another stark difference in the mechanics of peripheral oscillators versus those of the central circadian network, suggesting that the more classical theory of photic dominance holds true in the periphery.

Theoretically, cue dominance is related to the reliability of the sensory signal at conveying temporal information. One wonders if this property is innate, or something that might change with experience. Is it possible to devalue a cue? How much poor information must light provide before the clock ignores it altogether? These ideas could be explored in the laboratory, by, for example, manipulating the amplitude of Zeitgeber cycling. Alternatively, one might adopt a more evolutionary line of reasoning, investigating wild type genotypes isolated from different parts of the globe, where environmental signals conceivably carry more or less information. One could imagine that temperature is a less reliable timing cue near the equator, due to its relative constancy. An appreciation of how, and with what resolve, circadian clocks value Zeitgebers will be critical for a more holistic understanding of the circadian system, and may direct future interventions on clock function.

Deviating briefly from our discussion of clock input, we might also consider the output task of circadian systems, which is presumably one of organising biology for the time of day. In this thesis, I have made extensive use of locomotor activity as one such prominent output of the brain clock network. Peripheral oscillators, however, are believed to govern more local biology in the tissues in which they reside.

Expanding this idea, and linking it to previous findings of a proprioceptive chordotonal organ-dependent mechanical input pathway to the *Drosophila* clock (Simoni et al., 2014), I explored the potential for a peripheral oscillator in the fruit fly ear (a major chordotonal organ and a key mechanosensory apparatus). Here, relative quantification of clock gene transcripts showed robust circadian cycling in

tissue isolated from the second antennal segment (Chapter 7). These data provide early evidence for a peripheral clock in this tissue, raising further questions about its role and level of autonomy within the wider circadian system.

One could imagine that the task of a putative peripheral clock in the fly ear is to orchestrate circadian modulation of mechanosensation. This is indeed an exciting proposition, and if true, could provide a mechanism for clock-controlled gating of its own input. Future work is therefore anticipated to investigate functional changes in fruit fly mechanosensation across a 24 hr day. The importance of chordotonal organ signalling for entrainment to both temperature (Sehadova et al., 2009) and vibrational (Simoni et al., 2014) stimuli also hints at the possibility for the fly's own locomotor activity to provide a feedback signal to the clock. This would further blur the boundary between output and input of this system.

The experimental findings of this thesis provide deeper insight on the role of circadian clocks in *Drosophila melanogaster*. However, mathematical modelling allows a fully integrated scientific approach to understanding. The hidden Markov model developed in Chapter 8 goes some way toward addressing this need, representing the computational tasks put to the clock.

Much as multisensory investigations have typically been under-represented in chronobiology, so too are many of the mathematical techniques employed in Chapter 8. Drawing directly from the field of machine learning, my model was able to infer underlying phase from circadian datasets, and further captured statistical representations of the system's phase response dynamics. In this way, I hope to have shown the immediate potential for this more probabilistic approach to the study of circadian clocks.

Final Comment

My PhD research has never failed to intrigue and motivate me. Neuroscience is my passion, and I am grateful to have had the opportunity to explore it. In this thesis, I hope to have yielded deeper insight on circadian principles, and provided usable understanding of the multisensory capabilities of this system. The circadian clock is a fascinating model neuronal network; part localised (central), part distributed (pe-

ripheral), receiving multimodal input, and using this to drive complex coordinated output. It truly is a brain within a brain.

Appendix A

Chapter 6 Supplementary Data

A.1 Additional Bioluminescence Data

This data serves as a repeat experiment to validate the results presented in Section 6.3.3. The data was also used in the pooled quantifications presented in Fig. 6.7.

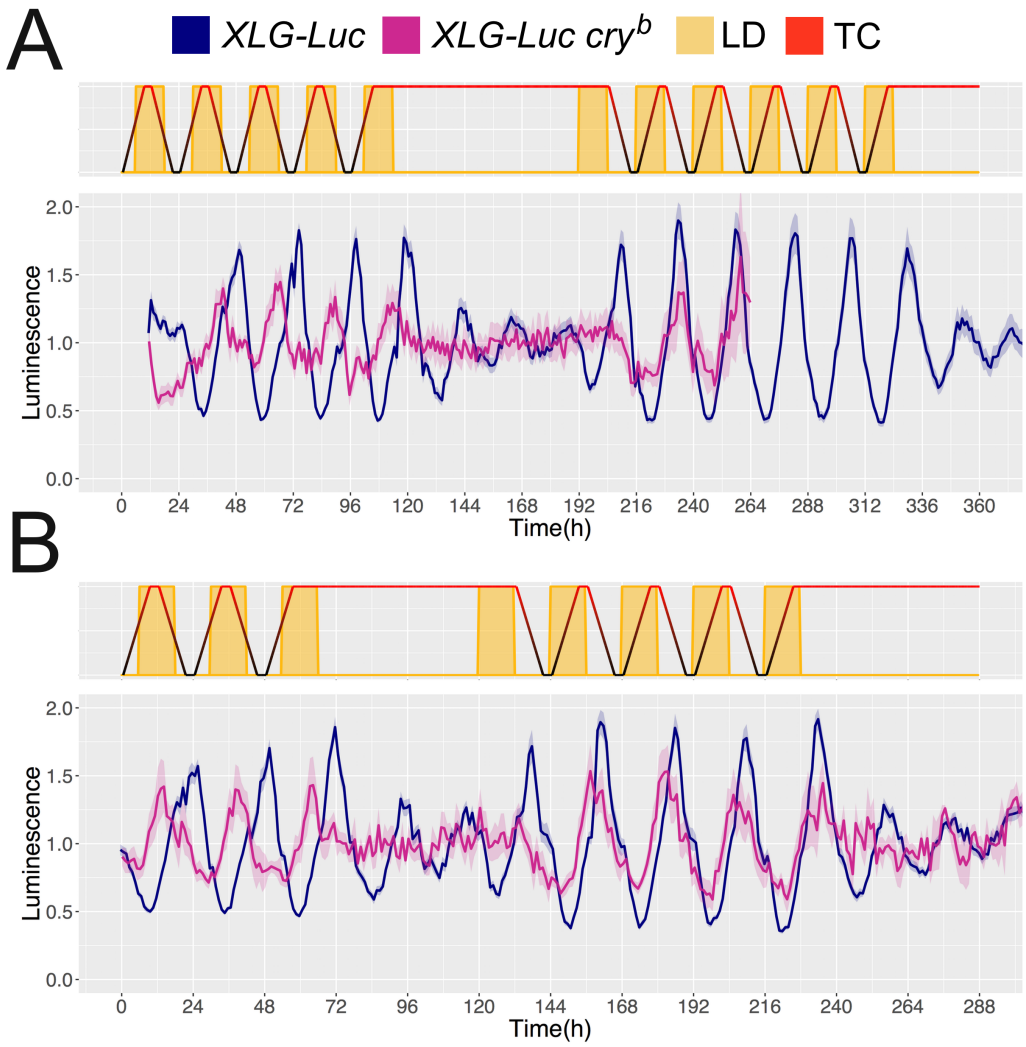

Figure A.1: Bioluminescence recordings of peripheral clocks during sensory conflict. (A) Experimental regime in which environmental conditions followed 5 days of misaligned LD:TC via 6 hr delay of LD, 2 days of free run in DD at 26°C, 6 days of aligned LD:TC, and finally 2 days of free run at DD and 26°C. (n = 41-44 and n = 8-19 for wild type and cry mutants respectively). *XLG-luc cry^b* flies died during the experiment, leading to incomplete data for this genotype. (B) A repeat of (A) where fewer days out-of-phase condition were used to improve chances of survival for the duration of the experiment (n = 43-48 and n = 7-10 for wild type and *cry* mutants respectively).

Appendix B

Chapter 8 Supplementary Data

B.1 Derivation of the HMM Objective Function

Maximum likelihood parameter learning in HMMs involves maximising a lower bound on the incomplete-data log-likelihood. We first introduce a density function over the hidden variable, $R(\mathbf{S})$, and through Jensen's inequality, we define our lower bound:

$$\begin{aligned}
\log \mathcal{L}(\lambda;\mathbf{O}) &= \log \sum_{\mathbf{S}} P(\mathbf{O},\mathbf{S};\lambda) \\
&= \log \sum_{\mathbf{S}} \frac{R(\mathbf{S})P(\mathbf{O},\mathbf{S};\lambda)}{R(\mathbf{S})} \\
&\geq \sum_{\mathbf{S}} R(\mathbf{S}) \log \frac{P(\mathbf{O},\mathbf{S};\lambda)}{R(\mathbf{S})} \\
&= E_{R(\mathbf{S})}[\log P(\mathbf{O},\mathbf{S};\lambda)] + H[R(\mathbf{S})] \\
&= L(R(\mathbf{S}),\lambda;\mathbf{O})
\end{aligned} \tag{B.1}$$

where $H[R(\mathbf{S})]$ is the entropy of the hidden variable distribution.

In this way, $L(R(\mathbf{S}),\lambda;\mathbf{O})$ is the objective function we seek to maximise. Starting with parameters λ^n, in the 'E-step', $L(R(\mathbf{S}),\lambda^n;\mathbf{O})$ is maximised with respect to the distribution $R(\mathbf{S})$ over the hidden variables; this sets $R^n(\mathbf{S}) = P(\mathbf{S}|\mathbf{O};\lambda^n)$. In the 'M-step', $L(R^n(\mathbf{S}),\lambda;\mathbf{O})$ is maximised with respect to the model parameters λ, leading to λ^{n+1}. This can be expressed in one step through definition of the $\mathcal{Q}$

function:

$$\mathcal{Q}(\lambda, \lambda^n) = E_{R^n(\mathbf{S})}[\log P(\mathbf{O}, \mathbf{S}; \lambda)] \tag{B.2}$$

where $R^n(\mathbf{S}) = P(\mathbf{S}|\mathbf{O}; \lambda^n)$ in the E-step.

B.2 Transition Probability Kernel Hyperparameter Learning

The standard parameter update rules for HMMs are derived by differentiating the objective function, setting this equal to zero, and solving for each parameter. Use of the transition probability kernel described in Eq. 8.5 required that learning occur for the μ and κ hyperparameters, rather than the probabilities a_{ij} directly (as occurs in the standard HMM framework).

We can write the $\mathcal{Q}$ function as:

$$\begin{aligned} \mathcal{Q}(\lambda, \lambda^n) &= E_{R^n(\mathbf{S})}[\log P(\mathbf{O}, \mathbf{S}; \lambda)] \\ &= \sum_{\mathbf{S}} \log P(\mathbf{O}, \mathbf{S}; \lambda) P(\mathbf{S}|\mathbf{O}; \lambda^n) \qquad \text{from the E step} \end{aligned} \tag{B.3}$$

Ignoring the Zeitgebers, the model defines

$$P(\mathbf{O}, \mathbf{S}; \lambda) = \pi_{S_1} b_{S_1 O_1} \prod_{t=2}^{T} a_{S_{t-1} S_t} b_{S_t O_t} \tag{B.4}$$

We can therefore separate the $\mathcal{Q}$ function into distinct terms for each parameter

$$\begin{aligned} \mathcal{Q}(\lambda, \lambda^n) = &\sum_{\mathbf{S}} \log \pi_{S_1} P(\mathbf{S}|\mathbf{O}; \lambda^n) + \sum_{\mathbf{S}} (\sum_{t=2}^{T} \log a_{S_{t-1} S_t}) P(\mathbf{S}|\mathbf{O}; \lambda^n)) \\ &+ \sum_{\mathbf{S}} (\sum_{t=1}^{T} \log b_{S_t O_t} P(\mathbf{S}|\mathbf{O}; \lambda^n) \end{aligned} \tag{B.5}$$

The transition probabilities are now separated into the second term, such that we need optimise only this term.

B.2.1 Differentiating the $\mathcal{Q}$ function with respect to μ and κ

Let us consider $\mu_{i_{adv}}$ as an example. Taking the derivative of the second term in Eq. B.5 with respect to $\mu_{i_{adv}}$

$$\begin{aligned}\frac{\partial \mathcal{Q}}{\partial \mu_{i_{adv}}} &= \frac{\partial}{\partial \mu_{i_{adv}}}\left(\sum_{\mathbf{S}}(\sum_{t=2}^{T}\log a_{S_{t-1}S_t})P(\mathbf{S}|\mathbf{O};\lambda^n))\right) \\ &= \frac{\partial}{\partial \mu_{i_{adv}}}\left(\sum_{k=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\log a_{kj}P(S_{t-1}=q_k,S_t=q_j|\mathbf{O},\lambda^n)\right) \\ &= \frac{\partial}{\partial \mu_{i_{adv}}}\left(\sum_{k=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\log a_{kj}\xi_t^n(k,j)\right)\end{aligned} \tag{B.6}$$

where $\xi_t^n(k,j)$ is calculated in Eq. 8.16, using parameters λ^n.

From Eq. 8.5, we have

$$a_{ij} = \frac{(1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}}{\sum_{j=1}^{N}\left((1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)} \tag{B.7}$$

And from Eq. 8.6, we have

$$\begin{aligned}\tau_{i_{del}} &= \mu_{i_{del}} + Z_t\kappa_{i_{del}} \\ \tau_{i_{adv}} &= \mu_{i_{adv}} + Z_t\kappa_{i_{adv}}\end{aligned} \tag{B.8}$$

Substituting Eq. B.7 and Eq. B.8 into Eq. B.6 and solving for $\mu_{i_{adv}}$ gives:

$$\begin{aligned}\frac{\partial \mathcal{Q}}{\partial \mu_{i_{adv}}} = \sum_{j=1}^{N}\sum_{t=2}^{T}\Bigg(&\frac{-2\tau_{i_{adv}}g(i,j)Ce^{-\tau_{i_{adv}}^2 g(i,j)}}{(1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}} \\ &- \frac{\sum_{j=1}^{N}\left(-2\tau_{i_{adv}}g(i,j)Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}{\sum_{j=1}^{N}\left((1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}\Bigg)\xi_t^n(i,j)\end{aligned} \tag{B.9}$$

Similarly, we can write the derivative of the $\mathcal{Q}$ function with respect to $\mu_{i_{del}}$,

$\kappa_{i_{adv}}$ and $\kappa_{i_{del}}$:

$$\frac{\partial \mathcal{Q}}{\partial \mu_{i_{del}}} = \sum_{j=1}^{N}\sum_{t=2}^{T}\left(\frac{-2\tau_{i_{del}} g(i,j)(1-C)e^{-\tau_{i_{del}}^2 g(i,j)}}{(1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}} - \frac{\sum_{j=1}^{N}\left(-2\tau_{i_{del}} g(i,j)(1-C)e^{-\tau_{i_{del}}^2 g(i,j)}\right)}{\sum_{j=1}^{N}\left((1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}\right)\xi_t^n(i,j) \tag{B.10}$$

$$\frac{\partial \mathcal{Q}}{\partial \kappa_{i_{adv}}} = \sum_{j=1}^{N}\sum_{t=2}^{T}\left(\frac{-2Z_t\tau_{i_{adv}} g(i,j)Ce^{-\tau_{i_{adv}}^2 g(i,j)}}{(1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}} - \frac{\sum_{j=1}^{N}\left(-2Z_t\tau_{i_{adv}} g(i,j)Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}{\sum_{j=1}^{N}\left((1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}\right)\xi_t^n(i,j) \tag{B.11}$$

$$\frac{\partial \mathcal{Q}}{\partial \kappa_{i_{del}}} = \sum_{j=1}^{N}\sum_{t=2}^{T}\left(\frac{-2Z_t\tau_{i_{del}} g(i,j)(1-C)e^{-\tau_{i_{del}}^2 g(i,j)}}{(1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}} - \frac{\sum_{j=1}^{N}\left(-2Z_t\tau_{i_{del}} g(i,j)(1-C)e^{-\tau_{i_{del}}^2 g(i,j)}\right)}{\sum_{j=1}^{N}\left((1-C)e^{-\tau_{i_{del}}^2 g(i,j)} + Ce^{-\tau_{i_{adv}}^2 g(i,j)}\right)}\right)\xi_t^n(i,j) \tag{B.12}$$

Notice that when $Z_t = 0$, $\frac{\partial \mathcal{Q}}{\partial \kappa_{i_{adv}}}$ and $\frac{\partial \mathcal{Q}}{\partial \kappa_{i_{del}}}$ evaluate to zero. This proves that no learning occurs for these parameters, apart from when $Z_t = 0$. This condition is true only for a limited number of data points in the time series, hence the need for an additional smoothness penalty (discussed in Section 8.3.5.4).

B.3 Dynamical System Model PRC

A fifth-order Fourier series was fit using the least squares procedure to emulate the data presented in Vinayak et al. (2013). A plot of this PRC can be found in Fig. 8.7, and its functional form was as follows:

$$\begin{aligned}\Phi(\theta_t) = \alpha_0 &+ \alpha_1\cos(\theta_t w) + \beta_1\sin(\theta_t w) + \alpha_2\cos(2\theta_t w) + \beta_2\sin(2\theta_t w) \\ &+ \alpha_3\cos(3\theta_t w) + \beta_3\sin(3\theta_t w) + \alpha_4\cos(4\theta_t w) + \beta_4\sin(4\theta_t w) \\ &+ \alpha_5\cos(5\theta_t w) + \beta_5\sin(5\theta_t w)\end{aligned} \tag{B.13}$$

Fitted parameter values shown in Table B.1.

Parameter	Value
α_0	-0.5305
α_1	1.2750
β_1	0.8044
α_2	0.4494
β_2	-1.3140
α_3	-0.6153
β_3	-0.3178
α_4	-0.2037
β_4	0.1030
α_5	-0.01818
β_5	0.07764
w	0.004363

Table B.1: Fitted PRC parameter values

B.4 Calculation of Performance Metric, $\mathcal{J}$

To quantify predictive performance of the statistical models presented in Chapter 8, metric $\mathcal{J}$ was calculated as follows:

$$\mathcal{J} = \frac{\sum_{t=1}^{T} \varepsilon_{HMM_t}}{\sum_{t=1}^{T} \varepsilon_{Naive_t}} \tag{B.14}$$

where ε_{HMM_t} and ε_{Naive_t} are the distance between predicted phase state and true phase value respectively. (See Fig. B.1).

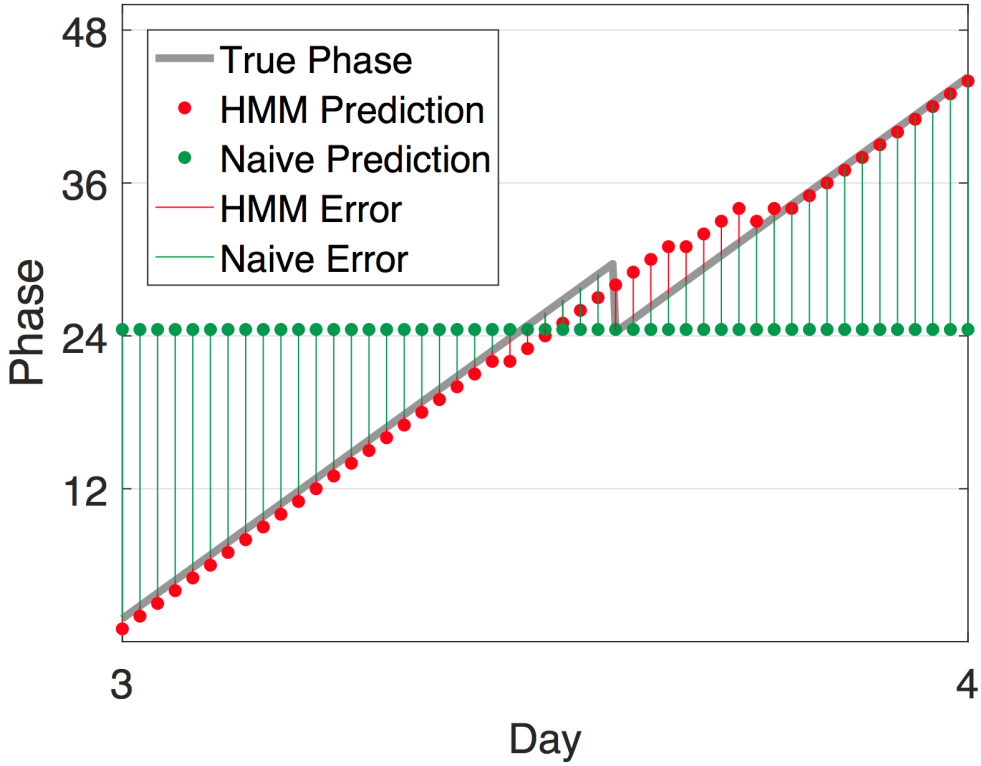

Figure B.1: Calculation of the $\mathcal{J}$ performance metric. The residual difference between model predictions (red) and the true phase values (grey) are calculated and summed. These are then divided by the summed residual difference between true phase values and the predictions of a naive model that simply estimates hidden phase to be at the midway point (12 hr) (green).

Appendix C

Colophon

This document was set in the Times Roman typeface using LaTeX and BibTeX, composed in the TeXstudio environment. Graphical drawings of the model architectures presented in Chapter 8 were typeset using the TikZ/PGF packages.

Bibliography

Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. *Chronobiology International*, 29(9):1153–1175.

Akten, B., Jauch, E., Genova, G. K., Kim, E. Y., Edery, I., Raabe, T., and Jackson, F. R. (2003). A role for CK2 in the Drosophila circadian oscillator. *Nature Neuroscience*, 6(3):251–257.

Albert, J. T. and Goepfert, M. C. (2015). Hearing in Drosophila. *Current Opinion in Neurobiology*, 34:79–85.

Albert, J. T., Nadrowski, B., and Göpfert, M. C. (2007). Mechanical signatures of transducer gating in the Drosophila ear. *Current Biology*, 17(11):1000–1006.

Allada, R., White, N. E., So, W. V., Hall, J. C., and Rosbash, M. (1998). A mutant Drosophila homolog of mammalian clock disrupts circadian rhythms and transcription of period and timeless. *Cell*, 93(5):791–804.

Bae, K., Lee, C., Sidote, D., Chuang, K. Y., and Edery, I. (1998). Circadian regulation of a Drosophila homolog of the mammalian Clock gene: PER and TIM function as positive regulators. *Molecular and cellular biology*, 18(10):6142–51.

Bass, J. (2012). Circadian topology of metabolism. *Nature*, 491(7424):348–56.

Batschelet, E. (1981). Circular statistics in biology. In *Circular Statistics in Biology*, pages 3–44.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.

Bengio, Y. and Frasconi, P. (1995). An input output HMM architecture. *Neural Information Processing Systems*, pages 427–434.

Bieler, J., Cannavo, R., Gustafson, K., Gobet, C., Gatfield, D., and Naef, F. (2014). Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular systems biology*, 10:739.

Boothroyd, C. E., Wijnen, H., Naef, F., Saez, L., and Young, M. W. (2007). Integration of light and temperature in the regulation of circadian gene expression in Drosophila. *PLoS Genetics*, 3(4):0492–0507.

Brandes, C., Plautz, J., and Stanewsky, R. (1996). Novel features of Drosophila period transcription revealed by real-time luciferase reporting. *Neuron*, 16:687–692.

Bretzl, H. (1903). Botanische Forschungen des Alexanderzuges. *Nature*, 68(1761):292–293.

Brown, E. N. and Czeisler, C. A. (1992). The statistical analysis of circadian phase and amplitude in constant-routine core-temperature data. *Journal of Biological Rhythms*, 7(3):177–202.

Brown, S. A., Kowalska, E., and Dallmann, R. (2012). (Re)inventing the circadian feedback loop. *Developmental Cell*, 22(3):477–487.

Caldwell, J. C. and Eberl, D. F. (2002). Towards a molecular understanding of Drosophila hearing.

Cavallari, N., Frigato, E., Vallone, D., Fröhlich, N., Lopez-Olmeda, J. F., Foà, A., Berti, R., Sánchez-Vázquez, F. J., Bertolucci, C., and Foulkes, N. S. (2011). A blind circadian clock in cavefish reveals that opsins mediate peripheral clock photoreception. *PLoS Biology*, 9(9):e1001142.

Collins, B., Mazzoni, E. O., Stanewsky, R., and Blau, J. (2006). Drosophila CRYPTOCHROME is a circadian transcriptional repressor. *Current Biology*, 16(5):441–449.

Correa, A., Lewis, Z. A., Greene, A. V., March, I. J., Gomer, R. H., and Bell-Pedersen, D. (2003). Multiple oscillators regulate circadian gene expression in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23):13597–602.

Covington, M. F. and Harmer, S. L. (2007). The circadian clock regulates auxin signaling and responses in Arabidopsis. *PLoS Biology*, 5(8):1773–1784.

Currie, J., Goda, T., and Wijnen, H. (2009). Selective entrainment of the Drosophila circadian clock to daily gradients in environmental temperature. *BMC biology*, 7:49.

Curtin, K. D., Huang, Z. J., and Rosbash, M. (1995). Temporally regulated nuclear entry of the Drosophila period protein contributes to the circadian clock. *Neuron*, 14(2):365–372.

Cyran, S. A., Buchsbaum, A. M., Reddy, K. L., Lin, M. C., Glossop, N. R. J., Hardin, P. E., Young, M. W., Storti, R. V., and Blau, J. (2003). vrille, Pdp1, and dClock form a second feedback loop in the Drosophila circadian clock. *Cell*, 112(3):329–341.

Darlington, T. K. (1998). Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science*, 280(5369):1599–1603.

De Mairan, J. (1729). Observation Botanique. *Histoire de l'Academie Royale des Sciences*, page 35.

Dekens, M. P. S., Santoriello, C., Vallone, D., Grassi, G., Whitmore, D., and Foulkes, N. S. (2003). Light regulates the cell cycle in zebrafish. *Current Biology*, 13(23):2051–2057.

Dempster, A., Laird, N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, 39(1):1–38.

Dolezelova, E., Dolezel, D., and Hall, J. C. (2007). Rhythm defects caused by newly engineered null mutations in Drosophila's cryptochrome gene. *Genetics*, 177(1):329–345.

Dong, W., Tang, X., Yu, Y., Nilsen, R., Kim, R., Griffith, J., Arnold, J., and Schüttler, H. B. (2008). System biology of the clock in Neurospora crassa. *PLoS ONE*, 3(8).

Drescher, K., Cornelius, G., and Rensing, L. (1982). Phase response curves obtained by perturbing different variables of a 24 hr model oscillator based on translational control. *Journal of Theoretical Biology*, 94(2):345–353.

Eastman, C. I. and Burgess, H. J. (2009). How to travel the world without jet lag.

Eckel-Mahan, K. and Sassone-Corsi, P. (2013). Metabolism and the circadian clock converge. *American Physiological Society*, 93(1):107–135.

Eddy, S. R. (2004). What is a hidden Markov model? *Nature biotechnology*, 22(10):1315–1316.

Emery, P., So, W. V., Kaneko, M., Hall, J. C., and Rosbash, M. (1998). CRY, a Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, 95(5):669–679.

Escola, S., Fontanini, A., Katz, D., and Paninski, L. (2011). Hidden Markov models for the stimulus-response relationships of multistate neural systems. *Neural computation*, 23(5):1071–132.

Field, L. H. and Matheson, T. (1998). Chordotonal organs of insects. *Advances in Insect Physiology*, 27(C).

Forger, D. B. and Peskin, C. S. (2003). A detailed predictive model of the mammalian circadian clock. *Proceedings of the National Academy of Sciences*, 100(25):14806–14811.

Fourier, J. (1822). *Théorie analytique de la chaleur*. Chez Firmin Didot.

Frisch, B., Hardin, P. E., Hamblen-Coyle, M. J., Rosbash, M., and Hall, J. C. (1994). A promoterless period gene mediates behavioral rhythmicity and cyclical per expression in a restricted subset of the Drosophila nervous system. *Neuron*, 12(3):555–570.

Froy, O. and Miskin, R. (2007). The interrelations among feeding, circadian rhythms and ageing. *Progress in Neurobiology*, 82(3):142–150.

Fujii, S., Krishnan, P., Hardin, P., and Amrein, H. (2007). Nocturnal male sex drive in Drosophila. *Current Biology*, 17(3):244–251.

Gentile, C., Sehadova, H., Simoni, A., Chen, C., and Stanewsky, R. (2013). Cryptochrome antagonizes synchronization of Drosophila's circadian clock to temperature cycles. *Current biology : CB*, 23(3):185–95.

Giebultowicz, J. and Hege, D. (1997). Circadian clock in Malpighian tubules. *Nature*, 386:664.

Giebultowicz, J., Ivanchenko, M., and Vollintine, T. (2001). *Organization of the insect circadian system: spatial and developmental expression of clock genes in peripheral tissues of Drosophila melanogaster*.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

Glaser, F. T. and Stanewsky, R. (2005). Temperature synchronization of the Drosophila circadian clock. *Current Biology*, 15(15):1352–1363.

Glossop, N. R. J., Houl, J. H., Zheng, H., Ng, F. S., Dudek, S. M., and Hardin, P. E. (2003). vrille feeds back to control circadian transcription of Clock in the Drosophila circadian oscillator. *Neuron*, 37(2):249–261.

Goepfert, M. C., Albert, J. T., Nadrowski, B., Kamikouchi, A., Göpfert, M. C., Albert, J. T., Nadrowski, B., and Kamikouchi, A. (2006). Specification of auditory sensitivity by Drosophila TRP channels. *Nature Neuroscience*, 9(8):999–1000.

Goldbeter, A. (1995). A Model for circadian oscillations in the Drosophila period protein (PER). *Proceedings of the Royal Society of London B: Biological Sciences*, 261(0962-8452):319—-324.

Gonze, D., Bernard, S., Waltermann, C., Kramer, A., and Herzel, H. (2005). Spontaneous synchronization of coupled circadian oscillators. *Biophysical Journal*, 89(1):120–129.

Goodwin, B. C. (1965). Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–437.

Göpfert, M. C., Humphris, A. D. L., Albert, J. T., Robert, D., and Hendrich, O. (2005). Power gain exhibited by motile mechanosensory neurons in Drosophila ears. *Proceedings of the National Academy of Sciences*, 102(2):325–30.

Göpfert, M. C. and Robert, D. (2003). Motion generation by Drosophila mechanosensory neurons. *Proceedings of the National Academy of Sciences*, 100(9):5514–5519.

Green, E. W., O 'callaghan, E. K., Hansen, C. N., Bastianello, S., Bhutani, S., Vanin, S., Armstrong, J. D., Costa, R., Kyriacou, C. P., and Hall, J. C. (2015). Drosophila circadian rhythms in seminatural environments: Summer afternoon component is not an artifact and requires TrpA1 channels. *Proceedings of the National Academy of Sciences*, 112(28):8702–8707.

Gu, C., Tang, M., and Yang, H. (2016). The synchronization of neuronal oscillators determined by the directed network structure of the suprachiasmatic nucleus under different photoperiods. *Scientific Reports*, 6(May):28878.

Gu, C., Wang, J., and Liu, Z. (2009). Free-running period of neurons in the suprachiasmatic nucleus: Its dependence on the distribution of neuronal coupling

strengths. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 80(3):5–8.

Gummadova, J. O., Coutts, G. A., and Glossop, N. R. J. (2009). Analysis of the Drosophila Clock promoter reveals heterogeneity in expression between subgroups of central oscillator cells and identifies a novel enhancer region. *Journal of Biological Rhythms*, 24(5):353–367.

Guo, F., Yu, J., Jung, H. J., Abruzzi, K. C., Luo, W., Griffith, L. C., and Rosbash, M. (2016). Circadian neuron feedback controls the Drosophila sleep-activity profile. *Nature*, 536(7616):292–297.

Halberg, F., Cornélissen, G., Katinas, G., Syutkina, E. V., Sothern, R. B., Zaslavskaya, R., Halberg, F., Watanabe, Y., Schwartzkopff, O., Otsuka, K., Tarquini, R., Frederico, P., and Siggelova, J. (2003). Transdisciplinary unifying implications of circadian findings in the 1950s. *Journal of circadian rhythms*, 1:2.

Halberg, F., Tong, Y. L., and Johnson, E. A. (1967). Circadian system phase - an aspect of temporal morphology; procedures and illustrative examples. *The Cellular Aspects of Biorhythms*, pages 20–48.

Hamasaka, Y., Rieger, D., Parmentier, M. L., Grau, Y., Helfrich-Förster, C., and Nässel, D. R. (2007). Glutamate and its metabotropic receptor in Drosophila clock neuron circuits. *Journal of Comparative Neurology*, 505(1):32–45.

Hardin, P. E. (1994). Analysis of period mRNA cycling in Drosophila head and body tissues indicates that body oscillators behave differently from head oscillators. *Molecular and cellular biology*, 14(11):7211–8.

Hardin, P. E., Hall, J. C., and Rosbash, M. (1990). Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels. *Nature*, 343(6258):536–540.

Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H. S., Han, B., Zhu, T., Wang, X., Kreps, J. a., and Kay, S. a. (2000). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science*, 290(5499):2110–2113.

Harper, R. E. F., Dayan, P., Albert, J. T., and Stanewsky, R. (2016). Sensory conflict disrupts activity of the Drosophila circadian network. *Cell Reports*, 17(7):1711–1718.

Harper, R. E. F., Ogueta, M., Dayan, P., Stanewsky, R., and Albert, J. T. (2017). Light dominates peripheral circadian oscillations in Drosophila melanogaster during sensory conflict. *Journal of Biological Rhythms*, 32(5).

Hastings, J. W. and Sweeney, B. M. (1958). A persistent diurnal rhythm of luminescence in Ganyaulax Polyedra. *The Biological Bulletin*, 115(3):440–458.

Helfrich-Forster, C. (1995). The period clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 92(2):612–616.

Helfrich-Förster, C. (2004). The circadian clock in the brain: A structural and functional comparison between mammals and insects.

Helfrich-Förster, C., Yoshii, T., Wülbeck, C., Grieshaber, E., Rieger, D., Bachleitner, W., Cusamano, P., and Rouyer, F. (2007). The lateral and dorsal neurons of Drosophila melanogaster: new insights about their morphology and function. *Cold Spring Harbor symposia on quantitative biology*, 72:517–25.

Hughes, M. E., DiTacchio, L., Hayes, K. R., Vollmers, C., Pulivarthy, S., Baggs, J. E., Panda, S., and Hogenesch, J. B. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genetics*, 5(4).

Hughes, M. E., Grant, G. R., Paquin, C., Qian, J., and Nitabach, M. N. (2012). Deep sequencing the circadian and diurnal transcriptome of Drosophila brain. *Genome Research*, 22(7):1266–1281.

Indic, P., Gurdziel, K., Kronauer, R. E., and Klerman, E. B. (2006). Development of a two-dimension manifold to represent high dimension mathematical models of the intracellular Mammalian circadian clock. *Journal of Biological Rhythms*, 21(3):222–232.

Indic, P., Schwartz, W. J., and Paydarfar, D. (2008). Design principles for phase-splitting behaviour of coupled cellular oscillators: clues from hamsters with 'split' circadian rhythms. *Journal of the Royal Society, Interface*, 5(25):873–883.

Ito, C., Goto, S. G., Shiga, S., Tomioka, K., and Numata, H. (2008). Peripheral circadian clock for the cuticle deposition rhythm in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24):8446–51.

Ito, C. and Tomioka, K. (2016). Heterogeneity of the peripheral circadian systems in Drosophila melanogaster: A review. *Frontiers in Physiology*, 7(JAN):1–7.

Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2009). Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proceedings of the National Academy of Sciences*, 106(33):14168–73.

Ivanchenko, M., Stanewsky, R., and Giebultowicz, J. M. (2001). Circadian photoreception in Drosophila: functions of cryptochrome in peripheral and central clocks. *Journal of Biological Rhythms*, 16(3):205–215.

Johnson, C. H. (1999). Forty years of PRCs–what have we learned? *Chronobiology international*, 16(6):711–43.

Kadener, S., Stoleru, D., McDonald, M., Nawathean, P., and Rosbash, M. (2007). Clockwork Orange is a transcriptional repressor and a new Drosophila circadian pacemaker component. *Genes and Development*, 21(13):1675–1686.

Kamikouchi, A., Inagaki, H. K., Effertz, T., Hendrich, O., Fiala, A., Göpfert, M. C., and Ito, K. (2009). The neural basis of Drosophila gravity-sensing and hearing. *Nature*, 458(7235):165–71.

Kaneko, M. and Hall, J. C. (2000). Neuroanatomy of cells expressing clock genes in Drosophila: Transgenic manipulation of the period and timeless genes to mark the perikarya of circadian pacemaker neurons and their projections. *Journal of Comparative Neurology*, 422(1):66–94.

Kelleher, F. C., Rao, A., and Maguire, A. (2014). Circadian molecular clocks and cancer. *Cancer Letters*, 342(1):9–18.

Keller, M., Mazuch, J., Abraham, U., Eom, G. D., Herzog, E. D., Volk, H.-D., Kramer, A., and Maier, B. (2009). A circadian clock in macrophages controls inflammatory immune responses. *Proceedings of the National Academy of Sciences*, 106(50):21407–12.

Kernan, M. J. (2007). Mechanotransduction and auditory transduction in Drosophila. *Pflügers Archiv - European Journal of Physiology*, 454(5):703–720.

Khabirova, E., Chen, K.-F., O'Neill, J. S., and Crowther, D. C. (2016). Flyglow: Single-fly observations of simultaneous molecular and behavioural circadian oscillations in controls and an Alzheimer's model. *Scientific Reports*, 6(May):33759.

Ki, Y., Ri, H., Lee, H., Yoo, E., Choe, J., and Lim, C. (2015). Warming up your ticktock: temperature-dependent regulation of circadian clocks. *The Neuroscientist*, 21(5):503–18.

Kim, J., Chung, Y. D., Park, D.-Y., Choi, S., Shin, D. W., Soh, H., Lee, H. W., Son, W., Yim, J., Park, C.-S., Kernan, M. J., and Kim, C. (2003). A TRPV family ion channel required for hearing in Drosophila. *Nature*, 424(6944):81–84.

Kloss, B., Rothenfluh, A., Young, M. W., and Saez, L. (2001). Phosphorylation

of PERIOD is influenced by cycling physical associations of DOUBLE-TIME, PERIOD, and TIMELESS in the Drosophila clock. *Neuron*, 30(3):699–706.

Koh, K. (2006). JETLAG resets the Drosophila circadian clock by promoting light-induced degradation of TIMELESS. *Science*, 312(5781):1809–1812.

Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K., and Takahashi, J. S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, 338(6105):349–54.

Konopka, R. J. and Benzer, S. (1971). Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 68(9):2112–2116.

Konopka, R. J., Pittendrigh, C., and Orr, D. (1989). Reciprocal behaviour associated with altered homeostasis and photosensitivity of Drosophila clock mutants. *Journal of Neurogenetics*, 21(4):243–52.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PloS one*, 2(9):e943.

Krishnan, B., Dryer, S. E., and Hardin, P. E. (1999). Circadian rhythms in olfactory responses of Drosophila melanogaster. *Nature*, 400(6742):375–378.

Krishnan, B., Levine, J. D., Lynch, M. K., Dowse, H. B., Funes, P., Hall, J. C., Hardin, P. E., and Dryer, S. E. (2001). A new role for cryptochrome in a Drosophila circadian oscillator. *Nature*, 411(6835):313–317.

Krishnan, P., Chatterjee, A., Tanoue, S., and Hardin, P. E. (2008). Spike Amplitude of Single-Unit Responses in Antennal Sensillae Is Controlled by the Drosophila Circadian Clock. *Current Biology*, 18(11):803–807.

Krupp, J. J., Kent, C., Billeter, J. C., Azanchi, R., So, A. K. C., Schonfeld, J. A., Smith, B. P., Lucas, C., and Levine, J. D. (2008). Social Experience Modifies Pheromone Expression and Mating Behavior in Male Drosophila melanogaster. *Current Biology*, 18(18):1373–1383.

Kwon, Y., Shen, W. L., Shim, H. S., and Montell, C. (2010). Fine thermotactic discrimination between the optimal and slightly cooler temperatures via a TRPV channel in chordotonal neurons. *J Neurosci*, 30(31):10465–10471.

Kyriacou, C. P. and Hastings, M. H. (2010). Circadian clocks: genes, sleep, and cognition. *Trends in Cognitive Sciences*, 14(6):259–267.

Kyriacou, C. P., Oldroyd, M., Wood, J., Sharp, M., and Hill, M. (1990). Clock mutations alter developmental timing in Drosophila. *Heredity*, 64(3):395–401.

Lee, G., Bahn, J. H., and Park, J. H. (2006). Sex- and clock-controlled expression of the neuropeptide F gene in Drosophila. *Proceedings of the National Academy of Sciences*, 103(33):12580–12585.

Leise, T. L. (2013). Wavelet analysis of circadian and ultradian behavioral rhythms. *Journal of Circadian Rhythms*, 11(1):5.

Leloup, J. C. and Goldbeter, A. (1998). A model for circadian rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins. *Journal of Biological Rhythms*, 13(1):70–87.

Leloup, J.-C. and Goldbeter, A. (2003). Toward a detailed computational model for the mammalian circadian clock. *Proceedings of the National Academy of Sciences*, 100(12):7051–6.

Levine, J., Funes, P., Dowse, H., and Hall, J. (2002a). Signal analysis of behavioral and molecular cycles. *BMC Neuroscience*, 3(1):1.

Levine, J. D., Funes, P., Dowse, H. B., and Hall, J. C. (2002b). Advanced analysis of a cryptochrome mutation's effects on the robustness and phase of molecular cycles in isolated peripheral tissues of Drosophila. *BMC neuroscience*, 3:5.

Levine, J. D., Funes, P., Dowse, H. B., and Hall, J. C. (2002c). Resetting the circadian clock by social experience in Drosophila melanogaster. *Science*, 298(5600):2010–2012.

Lin, J.-M., Kilman, V. L., Keegan, K., Paddock, B., Emery-Le, M., Rosbash, M., and Allada, R. (2002). A role for casein kinase 2[alpha] in the Drosophila circadian clock. *Nature*, 420(6917):816–20.

Lin, Y., Stormo, G. D., and Taghert, P. H. (2004). The neuropeptide pigment-dispersing factor coordinates pacemaker interactions in the Drosophila circadian system. *J Neurosci*, 24(36):7951–7957.

Liu, C., Weaver, D. R., Strogatz, S. H., and Reppert, S. M. (1997). Cellular construction of a circadian clock: Period determination in the suprachiasmatic nuclei. *Cell*, 91(6):855–860.

Liu, L., Li, Y., Wang, R., Yin, C., Dong, Q., Hing, H., Kim, C., and Welsh, M. J. (2007). Drosophila hygrosensation requires the TRP channels water witch and nanchung. *Nature*, 450(7167):294–298.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and. *Methods*, 25:402–408.

Locke, J. C., Westermark, P. O., Kramer, A., and Herzel, H. (2008). Global parameter search reveals design principles of the mammalian circadian clock. *BMC Systems Biology*, 2(1):22.

Lu, W., Meng, Q.-J., Tyler, N. J., Stokkan, K.-A., and Loudon, A. S. (2010). A circadian clock is not required in an arctic mammal. *Current Biology*, 20(6):533–537.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature genetics*, 40(4):476–83.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information.* W. H. Freeman.

Martinek, S., Inonog, S., Manoukian, A. S., and Young, M. W. (2001). A role for the segment polarity gene shaggy/GSK-3 in the Drosophila circadian clock. *Cell*, 105(6):769–779.

Matsuo, T. (2003). Control mechanism of the circadian clock for timing of cell division in vivo. *Science*, 302(5643):255–259.

Maury, E., Hong, H. K., and Bass, J. (2014). Circadian disruption in the pathogenesis of metabolic syndrome. *Diabetes and Metabolism*, 40(5):338–346.

Menet, J. S., Abruzzi, K. C., Desrochers, J., and Rosbash, M. (2010). Dynamic PER repression in the Drosophila circadian clock : from on- to off-DNA. *Genes & Development*, 24:358–367.

Miyasako, Y., Umezaki, Y., and Tomioka, K. (2007). Separate sets of cerebral clock neurons are responsible for light and temperature entrainment of Drosophila circadian locomotor rhythms. *Journal of Biological Rhythms*, 22(2):115–26.

Mohawk, J. A., Green, C. B., and Takahashi, J. S. (2012). Central and Peripheral Circadian Clocks in Mammals. *Annual Review of Neuroscience*, 35(1):445–462.

Morris, C. J., Purvis, T. E., Hu, K., and Scheer, F. A. J. L. (2016). Circadian misalignment increases cardiovascular disease risk factors in humans. *Proceedings of the National Academy of Sciences*, 113(10):E1402–11.

Murphy, K. (2012). *Machine learning: a probabilistic perspective*.

Nitabach, M. N., Blau, J., and Holmes, T. C. (2002). Electrical silencing of Drosophila pacemaker neurons stops the free-running circadian clock. *Cell*, 109(4):485–495.

Nitabach, M. N. and Taghert, P. H. (2008). Organization of the Drosophila circadian control circuit. *Current Biology*, 18(2):R84–93.

Okamura, H. (1999). Photic induction of mPer1 and mPer2 in cry-deficient mice lacking a biological clock. *Science*, 286(5449):2531–2534.

Pace-schott, E. F. and Hobson, J. A. (2002). The neurobiology of sleep: genetics, cellular physiology and subcortical networks. *Nature Reviews Neuroscience*, 3(August):591–605.

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, 109(3):307–320.

Park, J. H. and Hall, J. C. (1998). Isolation and chronobiological analysis of a neuropeptide pigment-dispersing factor gene in Drosophila melanogaster. *Journal of Biological Rhythms*, 13(3):219–228.

Peirson, S. N. and Foster, R. G. (2015). Sleep and circadian rhythm disruption in psychosis. In *Circadian Medicine*, pages 271–282. John Wiley & Sons, Inc, 10.1002/97 edition.

Percival, D. and Walden, A. (2000). *Wavelet methods for time series analysis*.

Peschel, N., Chen, K. F., Szabo, G., and Stanewsky, R. (2009). Light-dependent interactions between the Drosophila circadian clock factors cryptochrome, jetlag, and timeless. *Current Biology*, 19(3):241–247.

Peschel, N. and Helfrich-Förster, C. (2011). Setting the clock by nature: circadian rhythm in the fruitfly Drosophila melanogaster. *FEBS Letters*, 585(10):1435–1442.

Pittendrigh, C. S. and Bruce, V. G. (1957). An oscillator model for biological clocks. In *Rhythmic and Synthetic Processes in Growth*, pages 75–110. Princeton University Press.

Plautz, J. D. (1997). Independent photoreceptive circadian clocks throughout Drosophila. *Science*, 278(5343):1632–1635.

Plikus, M. V., Van Spyk, E. N., Pham, K., Geyfman, M., Kumar, V., Takahashi, J. S., and Andersen, B. (2015). The circadian clock in skin. *Journal of Biological Rhythms*, 30(3):163–182.

Price, J. L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., and Young, M. W. (1998). double-time is a novel Drosophila clock gene that regulates period protein accumulation. *Cell*, 94(1):83–95.

Price, T. S., Baggs, J. E., Curtis, A. M., Fitzgerald, G. A., and Hogenesch, J. B. (2008). WAVECLOCK: wavelet analysis of circadian oscillation. *Bioinformatics*, 24(23):2794–2795.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(January):4–16.

Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian processes for machine learning (adaptive computation and machine learning)*. The MIT Press.

Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C., and Taghert, P. H. (1999). A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in Drosophila. *Cell*, 99(7):791–802.

Rieger, D., Shafer, O. T., Tomioka, K., and Helfrich-Förster, C. (2006). Functional analysis of circadian pacemaker neurons in Drosophila melanogaster. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(9):2531–43.

Rodriguez, J., Tang, C.-H. A., Khodor, Y. L., Vodala, S., Menet, J. S., and Rosbash, M. (2013). Nascent-Seq analysis of Drosophila cycling gene expression. *Proceedings of the National Academy of Sciences*, 110(4):E275–84.

Roenneberg, T. and Merrow, M. (2016). The circadian clock and human health. *Current Biology*, 26(10):R432–R443.

Ruoff, P., Vinsjevik, M., Monnerjahn, C., and Rensing, L. (1999). The Goodwin oscillator: on the importance of degradation reactions in the circadian clock. *Journal of Biological Rhythms*, 14(6):469–479.

Rutila, J. E., Suri, V., Le, M., So, W. V., Rosbash, M., and Hall, J. C. (1998). Cycle is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of Drosophila period and timeless. *Cell*, 93(5):805–814.

Sayeed, O. and Benzer, S. (1996). Behavioral genetics of thermosensation and hygrosensation in Drosophila. *Proceedings of the National Academy of Sciences*, 93(12):6079–84.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y. J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., Liceiri, K., Tomancak, P., and A., C. (2012). Fiji: an open source platform for biological image analysis. *Nature Methods*, 9(7):676–682.

Schlichting, M. and Helfrich-Foerster, C. (2015). Photic entrainment in drosophila assessed by locomotor activity recordings. *Methods in Enzymology*, 552:105–123.

Schmidt, C., Collette, F., Cajochen, C., and Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cog Neuropsychol*, 24(7):755–789.

Sehadova, H., Glaser, F. T., Gentile, C., Simoni, A., Giesecke, A., Albert, J. T., and Stanewsky, R. (2009). Temperature entrainment of Drosophila's circadian clock

involves the gene nocte and signaling from peripheral sensory tissues to the brain. *Neuron*, 64(2):251–266.

Sehgal, A., Price, J. L., Man, B., and Young, M. W. (1994). Loss of circadian behavioral rhythms and per RNA oscillations in the Drosophila mutant timeless. *Science*, 263(5153):1603–1606.

Sehgal, A., Rothenfluh-Hilfiker, A., Hunter-Ensor, M., Chen, Y., Myers, M. P., and Young, M. W. (1995). Rhythmic expression of timeless: a basis for promoting circadian cycles in period gene autoregulation. *Science*, 270(5237):808–10.

Shafer, O. T., Helfrich-Förster, C., Renn, S. C. P., and Taghert, P. H. (2006). Reevaluation of Drosophila melanogaster's neuronal circadian pacemakers reveals new neuronal classes. *Journal of Comparative Neurology*, 498(2):180–193.

Simoni, A., Wolfgang, W., Topping, M. P., Kavlie, R. G., Stanewsky, R., and Albert, J. T. (2014). A mechanosensory pathway to the Drosophila circadian clock. *Science*, 343(6170):525–8.

Stanewsky, R., Frisch, B., Brandes, C., Hamblen-Coyle, M. J., Rosbash, M., and Hall, J. C. (1997a). Temporal and spatial expression patterns of transgenes containing increasing amounts of the Drosophila clock gene period and a lacZ reporter: mapping elements of the PER protein involved in circadian cycling. *The Journal of Neuroscience*, 17(2):676–696.

Stanewsky, R., Jamison, C. F., Plautz, J. D., Kay, S. a., and Hall, J. C. (1997b). Multiple circadian-regulated elements contribute to cycling period gene expression in Drosophila. *EMBO Journal*, 16(16):5006–5018.

Stanewsky, R., Kaneko, M., and Emery, P. (1998). The cryb Mutation Identifies Cryptochrome as a Circadian Photoreceptor in Drosophila. *Cell*, 95:681–692.

Stanewsky, R., Lynch, K. S., Brandes, C., and Hall, J. C. (2002). Mapping of elements involved in regulating normal temporal period and timeless RNA ex-

pression patterns in Drosophila melanogaster. *Journal of Biological Rhythms*, 17(4):293–306.

Storch, K.-F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., and Weitz, C. J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83.

Strogatz, S. H. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20.

Tanoue, S., Krishnan, P., Krishnan, B., Dryer, S. E., and Hardin, P. E. (2004). Circadian clocks in antennal neurons are necessary and sufficient for olfaction rhythms in Drosophila. *Current Biology*, 14:638–649.

Tataroglu, O., Zhao, X., Busza, A., Ling, J., O'Neill, J. S., and Emery, P. (2015). Calcium and SOL protease mediate temperature resetting of circadian clocks. *Cell*, 163(5):1214–1224.

Taylor, S. R., Doyle, F. J., and Petzold, L. R. (2008). Oscillator model reduction preserving the phase response: application to the circadian clock. *Biophysical journal*, 95(4):1658–1673.

Tevy, M. F., Giebultowicz, J., Pincus, Z., Mazzoccoli, G., and Vinciguerra, M. (2013). Aging signaling pathways and circadian clock-dependent metabolic derangements. *Trends in Endocrinology and Metabolism*, 24(5):229–237.

The Royal Society (2017). Machine learning: the power and promise of computers that learn by example. Technical report, The Royal Society.

Todi, S. V., Sharma, Y., and Eberl, D. F. (2004). Anatomical and molecular design of the Drosophila antenna as a flagellar auditory organ. *Microscopy Research and Technique*, 63(6):388–399.

van der Horst, G. T., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, a. P., van Leenen, D., Buijs, R., Bootsma,

D., Hoeijmakers, J. H., and Yasui, A. (1999). Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature*, 398(6728):627–30.

Van Swinderen, B. and Andretic, R. (2003). Arousal in Drosophila. *Behavioural Processes*, 64(2):133–144.

Vanin, S., Bhutani, S., Montelli, S., Menegazzi, P., Green, E. W., Pegoraro, M., Sandrelli, F., Costa, R., and Kyriacou, C. P. (2012). Unexpected features of Drosophila circadian behavioural rhythms under natural conditions. *Nature*, 484(7394):371–5.

Veleri, S., Brandes, C., Helfrich-Förster, C., Hall, J. C., and Stanewsky, R. (2003). A Self-Sustaining, Light-Entrainable Circadian Oscillator in the Drosophila Brain. *Current Biology*, 13(20):1758–1767.

Vijayan, V., Zuzow, R., and O'Shea, E. K. (2009). Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences*, 106(52):22564–8.

Vinayak, P., Coupar, J., Hughes, S. E., Fozdar, P., Kilby, J., Garren, E., Yoshii, T., and Hirsh, J. (2013). Exquisite light sensitivity of Drosophila melanogaster cryptochrome. *PLoS genetics*, 9(7):e1003615.

Vollmers, C., Schmitz, R. J., Nathanson, J., Yeo, G., Ecker, J. R., and Panda, S. (2012). Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metabolism*, 16(6):833–845.

Walker, R. G., Willingham, A. T., and Zuker, C. S. (2000). A Drosophila mechanosensory transduction channel. *Science*, 287(5461):2229–2234.

Wang, W., Barnaby, J. Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.-u., Fu, X.-D., and Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. *Nature*, 470(7332):110–4.

Wang, Y., Ke, C., and Brown, M. B. (2003). Shape-Invariant Modeling of Circadian Rhythms with Random Effects and Smoothing Spline ANOVA Decompositions. *Biometrics*, 59(4):804–812.

Weber, F., Zorn, D., Rademacher, C., and Hung, H. C. (2011). Post-translational timing mechanisms of the Drosophila circadian clock. *FEBS Letters*, 585(10):1443–1449.

Wheeler, D. a., Hamblen-Coyle, M. J., Dushay, M. S., and Hall, J. C. (1993). Behavior in light-dark cycles of Drosophila mutants that are arrhythmic, blind, or both. *Journal of Biological Rhythms*, 8(1):67–94.

Winfree, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology*, 16(1):15–42.

Yack, J. E. (2004). The structure and function of auditory chordotonal organs in insects. *Microscopy Research and Technique*, 63(6):315–337.

Yao, Z. and Shafer, O. T. (2014). The Drosophila circadian clock is a variably coupled network of multiple peptidergic units. *Science*, 343(6178):1516–20.

Yoshii, T., Hermann, C., and Helfrich-Förster, C. (2010). Cryptochrome-positive and -negative clock neurons in Drosophila entrain differentially to light and temperature. *Journal of Biological Rhythms*, 25(6):387–98.

Yoshii, T., Hermann-Luibl, C., and Helfrich-Föorster, C. (2016). Circadian light-input pathways in Drosophila. *Communicative and Integrative Biology*, 9(1):1–8.

Yoshii, T., Vanin, S., Costa, R., and Helfrich-Förster, C. (2009). Synergic entrainment of Drosophila's circadian clock by light and temperature. *Journal of Biological Rhythms*, 24(6):452–64.

Zeng, H., Qian, Z., Myers, M. P., and Rosbash, M. (1996). Alight entrainment mechanism for the Drosophila circadian clock. *Nature*, 380:129–135.

Zhang, W., Yan, Z., Jan, L. Y., and Jan, Y. N. (2013). Sound response mediated by the TRP channels NOMPC, NANCHUNG, and INACTIVE in chordotonal organs of Drosophila larvae. *Proceedings of the National Academy of Sciences*, 110(33):13612–13617.

Zhang, Y., Liu, Y., Bilodeau-Wentworth, D., Hardin, P. E., and Emery, P. (2010). Light and temperature control the contribution of specific DN1 neurons to Drosophila circadian behavior. *Current Biology*, 20(7):600–605.